THE RELATIONSHIP BETWEEN PROGRESS TESTS AND THE END OF COURSE ASSESSMENT TEST AT FOUNDATION LEVEL AT BILKENT UNIVERSITY SCHOOL OF ENGLISH LANGUAGE (SUSEL)

A THESIS PRESENTED BY
MÜGE AYŞE GENCER
TO THE INSTITUTE OF ECONOMICS AND SOCIAL SCIENCES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF ARTS
IN TEACHING ENGLISH AS A FOREIGN LANGUAGE

BILKENT UNIVERSITY
JULY 1999

THE RELATIONSHIP BETWEEN PROGRESS TESTS AND THE END OF COURSE ASSESSMENT TEST AT FOUNDATION LEVEL AT BILKENT UNIVERSITY SCHOOL OF ENGLISH LANGUAGE (BUSEL)

A THESIS PRESENTED BY

MÜGE AYŞE GENCER

TO THE INSTITUTE OF ECONOMICS AND SOCIAL SCIENCES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF MASTER OF ARTS

IN TEACHING ENGLISH AS A FOREIGN LANGUAGE

BILKENT UNIVERSITY

JULY 1999

BILKENT UNIVERSITY

INSTITUTE OF ECONOMICS AND SOCIAL SCIENCES

MA THESIS EXAMINATION RESULT FORM

July 31,1999

The examining committee appointed by the Institute of Economics and Social Sciences for the thesis examination of the MA TEFL student

Müge Ayşe Gencer

has read the thesis of the student.

The committee has decided that thesis of the student is satisfactory.

Thesis Title: The Relationship between Progress Tests and the End of Course Assessment Test at Foundation Level Bilkent University School Of English Language

Thesis Advisor: Dr. Necmi Akşit
Bilkent University, MA TEFL Program

Committee Members: Dr. Patricia N. Sullivan
Dr. William E. Snyder
Michele Rajotte
Bilkent University, MA TEFL Program

We certify that we have read this thesis and that in our combined opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts.

Dr. Necmi Akşit
(Advisor)

Dr. Patricia N. Sullivan
(Committee Member)

Dr. William E. Snyder
(Committee Member)

Michele Rajotte
(Committee Member)

Approved for the
Institute of Economics and Social Sciences

Ali Karaosmanoğlu
Director
Institute of Economics and Social Sciences

TABLE OF CONTENTS

ABSTRACT

Title: The Relationship between Progress Tests and the End of Course Assessment Test at Foundation Level at Bilkent University School of English Language

Author: Müge Ayşe Gencer

Thesis Chairperson: Dr. William E. Snyder
Bilkent University MA TEFL Program

Committee Members: Dr. Necmi Akşit
Bilkent University MA TEFL Program
Dr. Patricia N. Sullivan
Bilkent University MA TEFL Program
Michele Rajotte
Bilkent University MA TEFL Program

Testing has always been an important part of every teaching and learning experience. Tests can serve a number of purposes. One of the important ones is that tests can help identify the strengths and weaknesses of learning so that necessary help can be provided to learners.

The goals of this study were to investigate both the extent to which overall scores on progress tests were indicative of students' performance on the end of course assessment test and the extent to which each section on progress tests were indicative of performance on the end of course assessment test in Group 1,who studied 2 weeks at zero beginners and 14 weeks at beginners levels, and

in Group 2, who studied 16 weeks at beginners levels at Bilkent University School of English Language (BUSEL).

The subjects were 223 beginners level students. 98 of them were in Group 1 and 125 of them were in Group 2. The students in Group 1 consisted of only 4-year students whereas those in Group 2 included both 2-year and 4-year students. Group 2 was a heterogenous group.

The overall scores and breakdown scores of students in both groups were gathered and analyzed systematically using Pearson Product-Moment Correlation Coefficient.

The study revealed that all the correlations of overall scores on progress tests and on the the end of course assessment test were positive and statistically significant at $p<.001$ level in both Group 1 and Group 2. Group 1 students' overall scores on progress tests showed the highest correlations with the one on the end of course assessment test. The correlations ranged from .64 to .82 and they were all statistically significant at .001 level.

On the other hand, the breakdown of scores showed a variation. All of them had positive correlations and statistically significant between .001 and .10. In general, it was difficult to predict students' performance on different sections of the end of course assessment test. The correlation coefficients indicated

that students in Group 1 had the highest correlations both on overall scores and on the breakdown of scores.

CHAPTER 1 INTRODUCTION

Curriculum and testing affect students, teachers and, in fact, institutions very much. Hills (1976) stated that the basic principle involved in taking advantage of testing was to improve the organization of instruction. "The curriculum and the tests that are used to ascertain whether students are learning it [sic] must be coordinated for instruction to be most effective" (p. 267).

Since testing is an important part of every teaching and learning experience, well-made tests of English can be used in a number of ways to help students, as well as teachers. According to Madsen (1983),

> First of all, such [well-made] tests can help create positive attitudes toward class. In the interest of motivation of efficient instruction, teachers almost universally aim at providing positive classroom experiences for their students. There are some important ways that testing can contribute to this. One that applies in nearly every class is a sense of accomplishment... Tests of appropriate difficulty, announced well in advance and

> covering skills scheduled to be evaluated, can also contribute to a positive tone by demonstrating the spirit of fair play and consistency with course objectives. (p. 4)

A second way of supporting students is that well-made tests of English can help students to master the language. Learning can also be enhanced by students' growing awareness of the objectives and the areas of emphasis in the course. In addition, by diagnostic features, tests can foster learning. Madsen (1983) pointed out that "they [tests] confirm what each person has mastered, and they point up those language items needing further attention. Naturally, a better awareness of course objectives and personal language needs can help ... [the] students adjust their personal goals" (p. 4).

Third, teachers can use tests to diagnose their own efforts as well as those of their students (Carroll & Hall, 1985). As the teachers record the test scores, they might ask themselves several questions: "Are my lessons on the right level?" "What areas do we need more work on?" and " Which points need reviewing?"

Next, tests can provide insights into ways that the evaluation process can be improved (Valette, 1977). For example, "Did the test cause anxiety or resentment?" "Did the test results reflect accurately how my students have been performing in class, and in their assigned work?"

In short, students, teachers, and even administrators can benefit from tests by confirming progress that has been made, and showing how they can best redirect their future efforts (Valette, 1977). Good tests can aid learning.

Background of the Study

Carroll and Hall (1985) stated that tests, curricula, and programs should fully complement each other since learners see them as instruments of success or failure. Tests in particular have a dominating role in the curriculum. They can have not only a stimulating effect on teaching programs but also a "washback". Hughes (1997) defined "washback" as the impact of testing on learning and teaching. As Bachman (1998) pointed out, "'Good' tests will provide 'good' instructional practice and vice versa" (p. 11). Therefore, tests are of the utmost importance in the institutions.

Bilkent University is an English-medium university in Ankara, Turkey. Students registering at Bilkent University must take a proficiency test called Certificate of Proficiency in English (COPE) which has two stages. The first stage serves as a placement test for students who are placed in four lower levels: foundation 0 (FOU 0), foundation 1 (FOU 1), foundation 2 (FOU 2), and intermediate (INT). FOU-0 is real beginners, FOU-1 is false beginners and FOU-2 is pre-intermediate levels. The second stage of the placement test separates the ones who are successful on the test, and directly go to their departments from those who enter Bilkent University School of English Language (BUSEL) to study at the upper-intermediate (UP), and pre-faculty (PF) levels.

At the beginning of 1998-1999 academic year 348 students started FOU-1 level. These students studied 8 weeks in FOU-1 taking 3 progress tests and another 8 weeks in FOU 2 taking 3 more progress tests. These progress tests serve as formative tests/achievement tests. If students fulfil the course requirements, that is to say, 90 % attendance, Learning Portfolios handed in as required, and at least 60 % success rate covering the 6 progress tests and teacher assessment,

they are allowed to sit the end of course assessment. The end of course assessment functions as a summative test/final achievement test.

At foundation level, in BUSEL, students are required to obtain a grade of 60% to progress to the intermediate level. Although some students move on to the next level, some repeat the same course. During this process, "'How am I getting on?' is the question the student[s] ask. [They want] to know how much progress [they are] making in mastering [their] course" (Carroll & Hall, 1985, p. 108). What has been taught and learned is measured by progress tests indicating how far students have approached their target (Hughes, 1997; Valette, 1977).

In BUSEL, at foundation level, six progress tests, prepared by the Testing Unit, are given to students. The question types, and topics on these progress tests and the end of course assessment are generally in line with the course books, and the course objectives. However, sometimes, the students who perform well on the progress tests do not show the same or similar performance on the end of course assessment test or vice versa. The teachers and administrators as well as the students, value the students' performance on these achievement tests

because it is an indicator of not only the students' success, but also of the effectiveness of the organisation as a whole.

Statement of the Problem

Bilkent University School of English Language is an institution which gives tuition to both 2-year and 4-year students. 2-year students are vocational school students, whereas 4-year students are faculty students. They enter this university having performed very differently on the University Entrance Exam, extremely low and extremely high, respectively. Then, in BUSEL these students are placed according to their scores on the placement test administered by BUSEL. In foundation level, there are 3 progress tests administered periodically in an 8-week course. The purpose of these tests is to give feedback to students on their progress, to make slight adjustments in the instruction and to provide support to those who need it. BUSEL is interested in understanding the degree to which the performance on the progress tests can predict students' performance on the end of course assessment test so that administrators can take action to help students reach the required standards for advancement. However, they do not have the evidence.

Purpose of the Study

The purpose of the study is to find out the relationship between students' overall scores on progress tests and on the end of course assessment test, and the relationship between scores of each section on progress tests and on the end of course assessment test which student take at 2+14 week and 16-week foundation courses at BUSEL in 1998-1999 academic year. From this point onward, 2+14 week foundation course will be called Group 1 and 16-week foundation course will be named Group 2. In Group 1, students study 2 weeks at real beginners(fou-0), 6 weeks at beginners(fou-1) and 8 weeks at pre-intermediate(fou-2) levels. In Group 2, students study 8 weeks at beginners and another 8 weeks at pre-intermediate levels.

Significance of the Study

A wide range of individuals, such as teachers, administrators, the Head and members of the Curriculum and Testing Unit at BUSEL, future MA TEFL participants, and testers in other universities, can benefit from this research. First, teachers, administrators, and testers at BUSEL will have a chance to see/review how faculty and vocational school students at BUSEL perform in 16-week foundation level courses. As a result of

this study some changes, in question type(s), wording, or topic(s), might be made in the progress tests and the end of course assessment test or the same tests could be re-used. Second, since such a research study on testing at BUSEL has not been conducted before, it can bring a different perspective to the Curriculum and Testing Unit. Third, future MA TEFL participants can use it as a basis for their studies, either by continuing the same study from a different angle or applying it to different levels in BUSEL or at other universities. Finally, although testers in other universities do not have access to tests prepared and given by BUSEL, they may use this research study to structure their assessment system so that they can redesign and/or develop their own tests.

Research Question(s)

In the study, the research questions will be as follows:

1. To what extent are the overall scores of six progress tests indicative of students' performance on the end of course assessment test in Group 1 and in Group 2 at BUSEL?

   Sub-questions

a. To what extent are the overall scores of six progress tests indicative of 2-year students'

performance on the end of course assessment test in Group 2 at BUSEL?

b. To what extent are the overall scores of six progress tests indicative of 4-year students' performance on the end of course assessment test in Group 2 at BUSEL?

2. To what extent are the scores of each section on six progress tests indicative of students' performance on the end of course assessment test in Group 1 and in Group 2 at BUSEL?

Sub-questions:

a. To what extent are the scores of each section on six progress tests indicative of 2-year students' performance on the end of course assessment test in Group 2?

b. To what extent are the scores of each section on six progress tests indicative of 4-year students' performance on the end of course assessment test in Group 2?

CHAPTER 2 REVIEW OF THE LITERATURE

Introduction

In the past, a great number of tests have encouraged a tendency to separate testing from teaching. Both testing and teaching are so closely interrelated that it is actually impossible to work in either field without being constantly concerned with the other (Heaton, 1988, p. 5).

This chapter reviews the related literature on testing in the following order: (a) approaches to language testing, (b) purposes of assessment, (c) norm-referenced measurement and criterion-referenced measurement, (d) types of classroom testing, (e) basic qualities of good language tests.

Approaches to Language Testing

Language tests can be roughly classified in the following order: (1) the structuralist approach, (2) the integrative approach, and (3) the communicative approach (Heaton, 1988). They should not be considered as limited to certain periods in the development of language testing. A useful test will generally include features of several of these approaches. These approaches can be defined as follows:

The structuralist approach is guided by the view that language learning is mainly concerned with the

systematic acquisition of a set of habits. The learner's mastery of the separate elements of the target language, phonology, vocabulary, and grammar, is tested using words and sentences completely apart from any context on the grounds that a larger sample of language forms can be covered in the test in a comparatively short time. The skills of speaking, listening, reading, and writing are also separated from one another as much as possible because it is considered essential to test one thing at a time (Heaton, 1988).

The integrative approach involves the testing of language in context and is therefore concerned mainly with meaning and the total communicative effect of discourse. Integrative tests are often designed to assess the learner's ability to use two or more skills simultaneously (Heaton, 1988).

The communicative approach to language testing is sometimes linked to the integrative approach. Although both approaches emphasize the importance of the meaning of utterances rather than their form and structure, there are, nevertheless, fundamental differences between the two approaches. Communicative tests are primarily concerned with how language is used in communication. They aim to incorporate tasks as close as possible to those facing the students in real life. Language use,

which is concerned with how people actually use the language for a multitude of different purposes, is emphasized (Heaton, 1988).

At BUSEL, progress tests and the end of course assessment test are given to foundation level students at certain time intervals. These tests consist of reading, listening, grammar, vocabulary and writing. In general, the progress tests and the end of course assessment test at foundation level seem to be a combination of the integrative approach and the communicative approach as much as possible.

Bachman and Palmer (1996) claim,

> We should be able to bring about improvement in instructional practice through the use of tests which incorporate with the principles of effective teaching and learning (p. 34).

This is what is intended through the use of the progress tests and the end of course assessment test at BUSEL.

In brief, the tests aim at covering a representative sample of syllabus objectives, having more reliable grading across and between courses and providing a positive effect on learning and teaching through washback and feedback.

The cumulative nature of progress tests enables recycling in both teaching and learning. These tests contain a mixture of language, lexis, and sub-skills, rather than focusing on one skill.

Heaton (1988) believes:

> Language testing constantly involves making compromises between what is ideal and what is practicable in a certain situation. Nevertheless this should not be used as an excuse for writing and administering poor tests: whatever the constraints of the situation, it is important to maintain ideals and goals, constantly trying to devise a test which is valid and reliable as much as possible - and which has a useful backwash effect on the teaching and learning leading to the test (p. 24).

The Purposes of Assessment

"While trying to establish the worth of anything, hence to evaluate it, we need information and we need yardsticks against which to judge not only the information we require, but the information we receive. In education, where we are concerned with the worth of such things as curricula, teaching methods, and course materials, one significant source of information,

although not the only one, is the performance of those being taught - the students" (Harlen, 1978, p. 12).

Although many writers (Brown, 1996; Gronlund, 1985; Harlen, 1978; Heaton, 1988; Macintosch & Hale, 1976; Ooosterhof, 1994; Valette, 1977) define purposes differently, there is common thread of agreement in all of them.

Tests can serve several purposes. They may be constructed as devices to reinforce learning and to motivate the student or as a means of assessing the student's performance in the language. In the former case, as stated by Heaton (1988)," the tests are guided by the teaching that has taken place,...[whereas in the latter] case, the teaching is guided largely by the tests" (p. 5). In fact, testing is believed to be useful to increase student success (Bloom et al., 1971; Ebel, 1980; Natriello & Dornbush, 1984). Therefore, it can be used in the classroom evaluation to observe the extent to which learning outcomes are achieved (Gronlund, 1985). Namely, it is a systematic process to determine the degree of students' achievement during the instruction. It is very important to many facets of the school programs. It can directly contribute both to the teaching and learning process which is used in the class

and to a number of school uses (Carroll & Hall, 1985; Gronlund, 1985).

It is possible to classify and describe evaluation in many different ways with respect to the purpose. Testing and evaluation not only take an important role in the classroom instruction but also contribute to curriculum development.

Evaluation procedures can be categorized and explained in several forms. Gronlund (1985) presents the categories in terms of the evaluation of students performance in the following order:

> 1. determine …[students] performance at the beginning of the instruction (placement evaluation)
> 2. monitor learning process during instruction (formative evaluation)
> 3. diagnose learning difficulties during instruction (diagnostic evaluation)
> 4. evaluate achievement at the end of instruction (summative evaluation) (p. 11).

Formative evaluation can be used to provide feedback to both students and teachers. Furthermore, it supports successful learning and teaching. Where needed,

necessary support is provided both to the students and the teachers.

At BUSEL, progress tests serve the same purpose. One of the aims is to provide concrete and specific feedback to students. For example, instead of saying "You are poor at reading", teacher says "You are poor at reading main ideas". In other words, the tests give chance to see progress on not only each skill but also sub-skill.

Summative evaluation comes at the end of a course, "…it [summative evaluation] is designed to determine the extent to which the instructional objectives have been achieved" (Gronlund, 1985, p. 12). This type of evaluation provides information for judging the appropriateness of the course objectives and the effectiveness of the course (Oosterhof, 1994).

Like summative evaluation, the end of course assessment test at BUSEL is designed for the same purpose. It tests what has been taught throughout the whole course using the course objectives.

Language teachers are in the business of fostering achievement in the form of language learning. The purpose of most language programs is to maximize the possibilities for students to achieve a high degree of language learning (Brown, 1996). Teachers can find

themselves making achievement decisions sooner or later. " *Achievement decisions* are decisions about the amount of learning that students have done. Such decisions may involve who will be advanced to the next level of study or which students should graduate "(Brown, 1996, p. 14).

All these purposes should be considered similar.

Norm-Referenced Measurement and Criterion-Referenced Measurement

Within a language program, tests can function in various ways. There are mainly two categories. The first one is *norm-referenced* (NR) and the other one is *criterion-referenced* (CR). The former category helps administrators and teachers to make program-level decisions and the latter category helps teachers to make classroom-level decisions (Brown, 1996). Program-level decisions are proficiency and placement decisions whereas classroom-level decisions are diagnostic and achievement decisions.

There is a tendency to assume that comparisons must be made between individuals. This is known as *norm-referencing.*

Spolsky (1988) points out that the *norm-referenced test* is a discriminating test which aims to discover how much each student has benefited from the course and it

spreads out students as widely as possible in terms of their ability.

Unlike the *norm-referenced* test, a *criterion-referenced* test is usually used to measure well-defined and fairly specific objectives, which are related to a particular course, or program (Hills, 1976). The purpose of the CRT is to measure the amount of learning that a student has achieved on each objective. Students know in advance what types of questions, tasks, and content to expect for each objective (Brown, 1996).

Criterion-referenced assessment uses predetermined levels of performance, assessment being made in relation to objectives. This has the obvious advantage that the criteria can be pitched at any level, the primary concern being to ensure that as many students as possible reach the requisite level. Typically, it is used for guidance and diagnosis. Spolsky (1988) adds that a criterion-referenced test is a mastery test, designed to establish how many students have achieved a certain standard, or whether an individual student has performed a given a task. For example, formative and summative tests are criterion-referenced tests. They assess to what extent students have achieved course objectives during or at the end of the course.

Types of Classroom Testing

Criterion-referenced tests can be used for a variety of instructional purposes. The most important ones in this research study are formative testing and summative testing.

Formative tests

Formative tests are given periodically during the instruction to monitor students' learning progress and to provide continuous feedback to students and teachers. They reveal learning weaknesses in need of correction and encourage successful learning. They cover units, chapters, particular set of skills, tasks covered or practised during instruction. These tests are typically criterion-referenced tests (Finocchiaro & Sako, 1983).

At BUSEL, progress tests serve for the same purposes.

Summative Tests

They are given at the end of a course. The results can be used for evaluating the effectiveness of the instruction. They include test items with a wide range of difficulty. They include test items with a wide range of difficulty (Finocchiaro & Sako, 1983).

The end of course assessment test at BUSEL is administered for similar purposes.

Basic Qualities of Good Language Tests

A test, like any other type of instrument used to measure, should give the same results every time it measures, and should measure exactly what it is supposed to measure. In language testing, these terms are defined as *reliability* and *validity*. Reliability and validity are vital measurement qualities (Bachman and Palmer, 1996).

Reliability

In examining the meaningfulness of test scores, we are concerned with demonstrating that they are not unduly affected by factors other than the ability being tested" (Bachman, 1990, p. 25). If errors in measurement affect test scores, the scores will not be meaningful, and cannot supply the basis for valid interpretation and use (Bachman, 1990). Unfortunately, all examinations are subject to inaccuracies. While some measurement error is inevitable, it is possible to quantify and minimise the presence of measurement error (Henning, 1987).

Reliability is one of the important points in measurement. It is a quality of test scores, and a perfectly reliable score, or measure, would be one free from errors of measurement (American Psychological Association, 1985). This sort of accuracy is reflected in obtaining similar results when measurement is repeated with different instruments, by different students, and on

different occasions (Harris, 1969; Henning, 1987; Hughes, 1997). As Henning (1987) says "reliability is a measure of accuracy, consistency, dependability or fairness of scores resulting from administration of a particular examination"(p. 73), which is needed in all exams.

Deale (1975) defines reliability as consistency, meaning how far the test would give the same results if it could be done again by the same students under the same conditions. He points out that it is, of course, a theoretical definition since such conditions would be almost impossible to impose, and, therefore, a perfectly reliable test would be equally impossible to produce. "The factor variability, even if it is inevitable, needs to be reduced to an acceptable minimum, and to do this it is necessary to identify the principal sources of variability; these would seem to be:

- variations in performance of the student taking the test. These may stem from extraneous influences such as physical or mental or nervous conditions and anxiety and stress related to taking the test. Not much can be done to prevent these factors, but the teacher can take them into account when interpreting the results.
- variations in the test. The test can only

measure a small sample of a [student's] ability and a different sample could give a different result.

- variations in the marking. Except for objective tests, the marker's judgement can be as variable as the [student's] performance... Variations can occur for a variety of reasons: for example, the marker's standards being affected after marking a set of either very good or very bad scripts; or the teacher subconsciously being influenced by knowledge of the [student] whose work is being marked" (Heaton, 1988, p. 89).

Validity

Reliability is an important quality which should be monitored in tests; however, it is not itself sufficient for claiming that a test doing a good job. In fact, reliability is a pre-condition for validity, but not sufficient for purposes of judging overall test quality (Madsen, 1983). Validity must also be examined.

Brown (1996) defines validity "as the degree to which a test measures what it claims, or purports, to be measuring"(p. 231). To make sound decisions in

educational institutions, the development and use of valid tests are vital.

There are three main categories that exist for investigating the validity of a test: content validity, construct validity and criterion-related validity.

Content Validity

Testers should decide whether the test is a representative sample of the content of whatever the test is designed to measure in order to investigate the content validity. To have good content validity, a test must reflect both the content of the course and the balance in the teaching which led up to it.

Brown (1996) warns that " ...Once [testers] have established that a test has adequate content validity, they must immediately explore other kinds of validity arguments (construct or criterion-related) so that they can assess the validity of the test in terms related to the specific performances of the types of students for whom the test was designed in the first place" (p. 239).

Construct Validity

Hughes (1997) points out that " a test, part of a test, or a testing technique is said to have construct validity if it can be demonstrated that it measures just the ability which it is supposed to measure" (p. 26). The word 'construct' refers to any underlying ability

which is hypothesized in a theory of language ability. The tester conducts an experiment to investigate the degree to which the test is measuring the construct for which it was designed.

Criterion-related Validity

"Criterion-related validity is sometimes called concurrent or predictive validity. These terms are just variations... Concurrent validity is criterion-related validity but indicates that both measures were administered at about the same time... Predictive validity…[indicates] that two sets of numbers are collected at different times... In fact, the purpose of the test should logically be 'predictive' " (Brown, 1996, p. 248). Henning (1987) defines predictive validity as "… an indication of how well a test predicts intended performance" (p. 196). Besides, "…[It] indicates how well performance on a test correlates with performance on [another]"(Oosterhof, 1994, p. 60).

The usual procedure to compute predictive validity is to correlate statistically the two sets of scores and to report the degree of relationship between them by means of a *correlation coefficient.*

Anastasi (1961), Brown (1996; 1998), Cronbach (1964), and Davies (1984) emphasize that both content and construct validity and concurrent or predictive validity

are needed in the process of test validation. Combining validities strengthens overall validity.

To sum up, as Bachman (1996) states "... [reliability and validity] are the qualities that provide the major justification for using test scores ... as a basis for making inferences or decisions" (p. 19).

CHAPTER 3 METHODOLOGY

Introduction

The aim of this study is, in general, to investigate the relationship between six progress tests and the end of course assessment given to foundation (beginners) level students at Bilkent University School of English Language (BUSEL) in 1998-1999 academic year. More specifically, this study aimed at finding out both the extent to which the overall scores on progress tests were indicative of students' performance on the end of course assessment test in Group 1 and Group 2 at foundation levels at Bilkent University School of English Language, and the extent to which scores on sections on progress tests were indicative of performance on the same sections on the end of course assessment test as far as Group 1 and Group 2, and 2-year students and 4-year students in these courses are concerned.

Subjects

This study was carried out at Bilkent University School of English Language (BUSEL) in Ankara, Turkey. Bilkent University is an English-medium university; therefore, the students in its various departments have to have a good command of English in order to be successful in their studies. At the beginning of each academic year, the new entrants to the departments take

an English proficiency exam, prepared by the Curriculum and Testing Unit at BUSEL. Those who are successful in this exam become freshman students in their departments. The students who fail the exam are given a placement test to be classified into various levels: foundation 0 (FOU-0), foundation 1 (FOU-1), foundation 2 (FOU-2), intermediate (INT), upper-intermediate (UP) and pre-faculty (PF). In the fou-0 level are real beginners, in the fou-1 level false beginners, and in fou-2 level pre-intermediate students study.

The subjects in this study were chosen from false beginners (FOU-1) level. This group was divided into two groups as Group 1 and Group 2. The ones in the first group studied 2 weeks at real beginners (fou-0), 6 weeks at false beginners(fou-1) and 8 weeks at pre-intermediate (fou-2) levels. As mentioned before, the students in Group 1 were placed at this level according to their scores on the placement test. On the placement test since they did not score as well as the students in Group 2 did, they studied at real beginners level for 2 weeks. Then, they continued their study at fou-1 and fou-2 levels. This group consisted of only 4-year students, i.e., faculty students.

The second group studied 8 weeks at false beginners and 8 weeks at pre-intermediate levels. They performed

better than group 1 on the placement test; therefore, they did not study at real beginners level.

The test scores of 348 students were gathered. Then, the students who were repeating the course or did not sit at least one exam for reasons such as health, or late registration were excluded from the study as suggested by Brown (1996). The total number of the subjects was 223.

98 of them were in the first group. All the students in this study were 4-year students. There were 125 subjects in the second group. Seventy-one of them were 2-year students, that is to say, vocational school students. 4-year vocational school students were also included in this group. Fifty-four of them were 4-year students, in other words, faculty students.

Materials

In this study, initially, overall scores of six progress tests and of one the end of course assessment test were used. Next, breakdown scores of these progress tests and of the end of course assessment test were collected.

Both of Group 1 and Group 2 took 6 progress tests throughout the course. The progress tests started in the fourth week of the instruction and continued every other week except progress test 4. This progress test was administered in the twelfth week. The last progress test

was given in the sixteenth week. This was followed by the end of course assessment test in the seventeenth week. The end of course assessment test functions as a final achievement test.

The progress tests and the end of course assessment test contained a mixture of language, lexis and sub skills rather than focusing on one skill. During each 8-week course, three progress tests were given in the 4th, 6th and 8th weeks. The last block or the last two blocks on Monday afternoon was/were the exam time. Teachers and students were informed of the task types in the progress test the Wednesday before the exam to reduce the anxiety for the students but also to avoid constant practice of only one task type in the classroom. However, on the end of course assessment test, they were not informed since they were responsible for everything they had studied and since it was a summative test. The end of course assessment test was double scored to get reliable information whereas progress tests were marked only by class teachers.

Procedures

In this study, overall scores and breakdown scores of 223 students on seven tests were gathered. The breakdown scores of these progress tests were entered onto the computer to be analyzed using Pearson

Product-Moment Correlation. The collection of data was finished on May 29,1999. To estimate this relationship between progress test overall scores and the end of course assessment test overall scores, first, the researcher put students in two groups, Group 1 and Group 2. Then she divided these groups as 2-year students and 4-year students (see Fig. 1).

Group 1	2 weeks fou-0 6 weeks fou-1 8 weeks fou-2	98 students	4-year (98 students)
Group 2	8 weeks fou-1 8 weeks fou-2	125 students	4-year (54 students) 2-year (71 students)

Fig.1: How the subjects were put into groups

In the next stage, she computed the correlation between the overall scores on each of the six progress tests and the scores on the end of course assessment on the basis of Group 1 and Group 2, and 4-year and 2-year students (see Fig. 2).

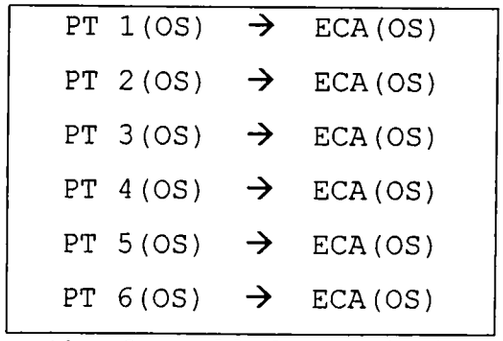

Fig. 2: How the Overall Scores(OS) were Correlated
Note. PT: Progress Test
ECA: the End of Course Assessment Test

The next step was the computation of breakdown scores of progress tests and of the end of course assessment test in Group 1, Group 2,and 4-year and 2-year students in these groups (see Figure 3).

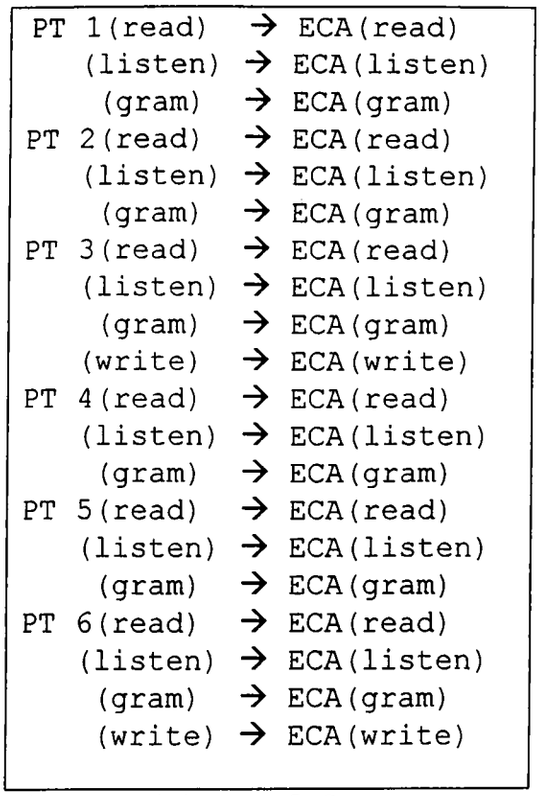

Fig. 3: How the Breakdown of Scores were Correlated
Note. PT: progress test
ECA: the end of course assessment test
read: reading listen: listening
gram: grammar write: writing

Data Analysis

After taking permission from BUSEL, the data were gathered. The 223 students' scores on six progress tests and one end of course assessment were analyzed to find the mean, the standard deviation and the correlation coefficient in terms of the type of course and the length of the study in departments. The Pearson Product-Moment Correlation was used to calculate the correlation coefficient.

After performing the necessary statistical techniques, the researcher used tables and figures to illustrate what she had discovered.

CHAPTER 4 DATA ANALYSIS

> A foolish man always thinks only of the results, and is impatient, without the effort that is necessary to get good results. No good can be attained without proper effort, just as there can be no third story (on a house) without the foundation and the first and second stories.
>
> (The teaching of the Buddha)

Overview of the Study

The purpose of this study was to find out both the extent to which the overall scores of progress tests were indicative of foundation level students' performance on the end of course assessment test at Bilkent University School of English Language; and the extent to which the overall scores of each section on each progress test were indicative of similar performance on the end of course assessment test with respect to Group 1 and Group 2, and 2-year students and 4-year students in Group 2.

The researcher obtained the six progress test scores and an end of course assessment test scores of 223 students. Then, she correlated the overall scores of

progress tests and the end of course assessment test for each group. Following this, she computed the correlation of each section on progress tests and the end of course assessment test for different groups in the study.

Data Analysis Procedures

This study investigated the degree of relationship between progress tests and the end of course assessment test, and the degree of relationship between each section on progress tests and on the end of course assessment test on the basis of Group 1 and Group 2.

This process was divided into several stages. In the first stage of data analysis, the researcher correlated progress test scores and the end of course assessment test scores using the Pearson Product Moment Correlation to see the degree of association between the progress tests and the end of course assessment test.

In the next stage of the analysis, again using the Pearson Product Moment Correlation Coefficient, each section on each progress test was correlated with each section on the end of course assessment test. To illustrate, reading section on each progress test was correlated with reading section on the end of course assessment test. Vocabulary was the only section which was not correlated since there was no vocabulary section on the end of course assessment test. Writing was

correlated only twice as it was tested only on the third and sixth progress tests. In addition, the central tendency and dispersion of each progress test and the end of course assessment test were computed to understand the anomalies that occurred in the correlation of these tests.

In the next stage, because of the fact that vocabulary was tested on five progress tests, but not on the end of course assessment test and writing was tested only on two progress tests, the researcher excluded vocabulary and writing sections on five progress tests and the end of course assessment test. Progress test 5 was the only test that tested neither vocabulary nor writing, but reading, listening and grammar.

In the end, there were only three sections on progress tests and on the end of course assessment test, which were reading, listening and grammar. The researcher correlated the overall scores on the progress test and on the end of course assessment test to see whether vocabulary and/or writing were the sources of differing correlations between progress tests and the end of course assessment test, using Pearson Product Moment Correlation Coefficient. The results of this calculation were interpreted after the correlation of overall scores.

Results

This section is divided into two main headings: correlation between overall scores on progress tests and the end of course assessment test (research question 1), and correlation between each section on progress tests and the end of course assessment test(research question 2). Under each section, sub-questions are also analyzed.

In the interpretation stage, the results of correlations are presented in an order of general results and specific results.

The correlation coefficients are interpreted with respect to the strength of the relationship, the direction of the relationship and the statistical significance of the correlation. As correlation coefficients range between -1.00 and +1.00, the strength and direction of the correlation are interpreted using the following values in Figure 4, suggested by Fitz~Gibbon and Morris (1987).

	+1.00	perfect positive correlation
+.99	+.80	very strong positive correlation
+.79	+.60	strong positive correlation
+.59	+.40	moderate positive correlation
+.39	+.20	weak correlation
+.19	-.20	no correlation
-.21	-.40	weak negative correlation
-.41	-.60	moderate negative correlation
-.61	-.80	strong negative correlation
-.81	-.99	very strong negative correlation
	-1.00	perfect negative correlation

Fig. 4: The Range of Possible Correlations and their usual Interpretations
Source: Fitz~Gibbon and Morris (82).

Research Question 1 : Correlation between Overall Scores on Progress Tests and on the End of Course Assessment Test

The degree of relationship between overall scores of progress tests and the end of course assessment test which was given to Group 1 and Group 2 at foundation level is interpreted in terms of the strength, the direction and the statistical significance of the correlation.

Group 1. This group is composed of 98 students, all of whom are 4-year students. Namely, they are faculty students. They studied 2 weeks at fou-0 and 14 weeks at fou-1 and fou-2.

Table 1
Correlation between Overall Scores of Progress Tests(PTs) and of the End of Course Assessment Test (ECA) in Group 1

n=98	ECA (week 17)
PT 1 (week 1)	.64 *
PT 2 (week 2)	.74 *
PT 3 (week 3)	.80 *
PT 4 (week 12)	.82 *
PT 5 (week 14)	.82 *
PT 6 (week 16)	.80 *

Note. n = number of students
* p< .001

To start with, the correlation between progress test 1 and the end of course assessment test is at .64. (see Table 1). This correlation is higher than expected for the first progress test because it is the first the

students sit for a progress test at BUSEL and it is just the beginning of the course. The correlation rises over the next four progress tests. It remains stable on progress test 5, but there is a slight decrease on progress test 6 (see Appendix A, Fig. 5).

The correlation between each progress test and the end of course assessment test is .64. This suggests that there is a strong positive relationship between progress test 1 and the end of course assessment test. This relationship is statistically significant at the .001 level. On progress test 2, the correlation coefficient is .74, which indicates a strong positive relationship between progress test 2 and the end of course assessment test. r is statistically significant at $p < .001$. $r = .80$ and $r = .82$ values imply that the correlation between progress tests 3, 4, 5 and the end of course assessment test is strongly positive. Regarding the degree of association, the closer to 1.00 in either direction, the greater the strength of correlation (Brown, 1996). The relationships are significant at the .001 level.

During the analysis process, the researcher notices that r decreases when progress test 6 is concerned. Thinking that vocabulary and writing might be the source of low correlation because of the fact that they are either tested very rarely or not tested on the end of

course assessment test, the researcher excludes vocabulary and writing from progress tests and the end of course assessment test. After this, overall scores on progress tests and on the end of course assessment test are correlated. However, it is found out that all the correlations between progress tests and the end of course assessment test further decrease apart from progress test 2. There is a slight increase in the correlation between overall score on progress test 2 and the end of course assessment test when writing on the end of course assessment test is not included (see Appendix C).

Group 2. There are 125 students in this group. On the placement test administered at the beginning of the academic year, these students scored better than the students in Group 1. Therefore, they did not study at fou-0 level as real beginners. 71 of these students are 2-year students and the rest are 4-year students.

Table 2
Correlation between Overall Scores of PTs and of ECA in Group 2

n=125	ECA (week 17)
PT 1 (week 4)	.54 *
PT 2 (week 6)	.66 *
PT 3 (week 8)	.73 *
PT 4 (week 12)	.74 *
PT 5 (week 14)	.80 *
PT 6 (week 16)	.63 *

Note. n = number of students
* p< .001

The correlation coefficient between progress tests and the end of course assessment test starts at .54 and increases steadily until it reaches progress test 5(see Appendix A, Fig. 6). On progress test 6, however, the correlation coefficient drops slightly as observed in Group 1. As Table 2 indicates, the highest correlation, r = .80, is between progress test 5 and the end of course assessment test. The lowest correlation is between progress test 1 and the end of course assessment test again.

There is a moderate positive correlation between progress test 1 and the end of course assessment test, with a value of .54 (see Table 2). It is also statistically significant at .001 level. The correlation coefficients between progress test 2, progress test 3,progress test 4 and the end of course assessment test range from .66 to .74, which means a strong positive correlation. These values have statistical significance at the .001 level. Namely, there is only .001% probability that these correlation coefficients occurred by chance. On the other hand, r = .80 suggests that there is a very strong positive and statistically significant relationship between overall scores on progress test 5 and the end of course assessment test. However, this correlation decreases to .63 as far as progress test 6 is

concerned. There is again a moderate positive and statistically significant correlation, but this value is not as high as the one between progress test 5 and the end of course assessment test.

When vocabulary and writing are not included in the overall scores on progress tests and the end of course assessment test, the correlation coefficients show a fall in values (see Appendix C).

Sub-question 1. Correlation between Overall Scores of PTs and ECA in 2-Year Students in Group 2

As Group 2 is a heterogeneous group, it is divided into two according to the departments the students will study in. These two groups are the 2-year student and 4-year student groups.

Table 3
Correlation between Overall Scores of 2-Year and 4-Year Students in Group 2 on PTs and on ECA

	ECA (week 17) (2-year Students) n=71	ECA (week 17) (4-year Students) n=54
PT 1 (week 4)	.52 *	.59*
PT 2 (week 6)	.63 *	.71*
PT 3 (week 8)	.70 *	.76*
PT 4 (week 12)	.65 *	.83*
PT 5 (week 14)	.73 *	.84*
PT 6 (week 16)	.55 *	.69*

Note. n = number of students
* $p < .001$

In the 2-year student group, as illustrated in Table 3, the correlation between progress tests and the

end of course assessment test starts at a moderate positive value and it goes up steadily. However, when it comes to progress test 6, there is a sudden decline (see Appendix A, Fig. 7).

The correlation between progress test 1, 6 and the end of course assessment test is moderately positive. The overall scores on progress test 2, 3, 4 and 5 have strong positive correlations with the overall scores on the end of course assessment test. All the values are also statistically significant at $p < .001$.

As mentioned in other groups, when vocabulary and writing are excluded from progress tests and the end of course assessment test, it is observed that r does not increase, which can be interpreted as vocabulary and writing are not the source of low correlations between progress tests and the end of course assessment test (see Appendix C).

Sub-question 2: Correlation between Overall Scores of Progress Tests and on the End of Course Assessment Test in 4-Year Students Group in Group 2

4-year Students

As Table 3 demonstrates, in 4-year student group, there is a gradual increase in the correlation coefficient except progress test 6. As far as the relationship between progress test 6 and the end of

course assessment test is concerned, there is a slight drop (see Appendix A Fig. 8).

Although r = .59 is a moderate positive correlation, it is also very close to r = .60, which would be a strong positive correlation. The relationship between progress test 1 and the end of course assessment test is moderately positive. Progress test 2, 3 and 6 have strong positive correlations with the end of course assessment test. Progress test 4 and progress test 5 have very strong positive relationships with the end of course assessment test. All the correlations are statistically significant at .001 level.

It is also found that when vocabulary and writing are not included in progress tests and the end of course assessment test, the correlation coefficient of the overall scores drops just as in the other groups (see Appendix C).

Research Question 2: Correlation between Each Section of Progress Tests and the End of Course Assessment Test

In this section, each section on progress tests is correlated with the one on the end of course assessment test. That is to say, the reading section is correlated with reading section or listening section is correlated with listening section. Sections are categorised into 4 groups depending on the sections on the end of course

assessment test as reading, listening, grammar and writing. Vocabulary is not included since it is not tested on the end of course assessment test. As described at the beginning of this chapter and shown in Table 4, 5 and 6, the breakdown scores of each progress test and of the end of course assessment test are correlated both for Group 1 and Group 2.

Group 1

Table 4
Correlation between Each Section of PTs the and ECA in Group 1

n=98	READING	LISTENING	GRAMMAR	WRITING
	ECA	ECA	ECA	ECA
PT 1	.36**	.46*	.61*	
PT 2	.23***	.31**	.71*	
PT 3	.30**	.41*	.69*	.49*
PT 4	.42*	74*	.54*	
PT 5	.53*	.37*	.76*	
PT 6	.47*	.69*	.45*	.55*

Note. n = number of students
* p< .001 ** p< .01 *** p< .05

In the reading section, the correlation coefficient fluctuates. It sometimes increases sometimes decreases.

Just like the correlation between overall scores on progress tests and on the end of course assessment test, r falls on progress test 6 (see Appendix Fig. 9).Progress test1, 2 and 3 have weak positive correlation with the end of course assessment test whereas progress test 4, 5, and 6 have moderate positive correlation. The statistical significance of these values vary from $p< .001$ to $p< .05$.

In the listening section, r starts at a .46 value. Just like the one in the reading section, it fluctuates. There is a sudden rise and decline. For example, r rises from .41 to .74 then declines to .37 followed by a rise to .69. In short, there is not a consistent increase or decrease as illustrated in Appendix A Fig.10. progress test 2 and 5 have weak positive and significant relationships with the end of course assessment test while progress test 1 and 3 have moderate positive correlations. Finally, r=.74 and r=.69 suggest that progress test 4and 6 have strong positive relationships with the end of course assessment test. They are statistically significant at .001 level.

Although correlation coefficient in grammar section fluctuates, it is relatively more consistent than reading and listening sections (see Appendix A Fig. 11). It starts at a higher value than other sections. Only

progress test 4 and 6 have moderate positive correlation, the rest have strong positive correlation. All of them are statistically significant at .001 level. Grammar is the only section where X is high when compared to other sections (see Appendix C). What is more striking is that although X of progress test 6 is the highest and SD is the lowest among all progress tests, the lowest correlation is between progress test 6 and the end of course assessment test in grammar section.

As far as the writing section is concerned, it is really difficult to make any estimations since writing is tested only twice on progress tests and using different criteria on each of them (see Appendix A Fig.12). In addition, there is a long interval between progress test 3, which is administered in week 6, and the end of course assessment test, which is administered in week 17. When looked at separately, it seems that there is a moderate positive correlation between writing section on progress tests and the end of course assessment test. Both r values are statistically significant.

To sum up, sections on progress tests do not show very strong positive correlations with the end of course assessment test. Therefore, it is difficult to make estimations about students' performance on different sections on the end of course assessment test.

Group 2

Table 5
Correlation between Each Section of PTs and on ECA in Group 1

n=125	READING	LISTENING	GRAMMAR	WRITING
	ECA	ECA	ECA	ECA
PT 1	.17*****	.19****	.51*	
PT 2	.44*	.39*	.62*	
PT 3	.19****	.23***	.58*	.47*
PT 4	.36*	.53*	.52*	
PT 5	.44*	.27**	.68*	
PT 6	.13	.57*	.55*	.34*

Note. n = number of students
* p< .001 *** p< .02 ***** p< .10
** p< .01 **** p< .05

In the reading section, the correlation coefficient starts at .17 and then it fluctuates. There are steep rises and declines as illustrated in Appendix A Fig.13. It is observed that there is no correlation between progress test 1, 3, 6 and the end of course assessment test. It is supposed that progress test 6 is the one which is the most similar to the end of course assessment test in terms of content and is the closest one in time

since progress test 6 is administered in week 16 and the end of course assessment test is administered in week 17. On the contrary, as far as reading section is concerned, the correlation coefficient between progress test 6 and the end of course assessment test is the lowest one among the others. All r values are statistically significant except progress test 6.

The correlation coefficient goes from .19 to .57 in the listening section. However, there is not a sustained rise but abrupt plunges (see Appendix A Fig.14). The statistical significance of r varies between .05 and .001

The r in grammar section values higher than the ones in the other sections. There is fluctuation among correlation coefficients but it is not as steep as the ones in the reading and listening sections (see Appendix A Fig. 15). Progress test 1, 3, 4, and 6 have moderate positive relationships with the end of course assessment test. Progress test 2 and 5 have strong positive correlation with the end of course assessment test. All of the correlation coefficients have significance at .001 level. It seems that it is easier to estimate Group 2 students' performance on grammar section on the end of course assessment test than other sections.

The correlation coefficient in the writing section starts at .47 suggesting moderate positive relationship

then this goes down .34, which means weak positive relationship with the end of course assessment test. These values are statistically significant at .001 level. In fact, it is almost impossible to say anything about students' probable performance on writing on the end of course assessment test (see Appendix A Fig. 16).

Sub-question 1: Correlation between Each Section on PTs and ECA in 2-Year Students Group in Group 2

2-Year Students Group in Group 2

Table 6
Correlation between Each Section on PTs and ECA in 2-Year Students in Group 2

n=71 (2-year students)	READING ECA	LISTENING ECA	GRAMMAR ECA	WRITING ECA
PT 1	.00	.15	.46*	
PT 2	.35**	.44*	.57*	
PT 3	.17	.15	.60*	.42*
PT 4	.33**	.48*	.59*	
PT 5	.41*	.26****	.59*	
PT 6	.10	.60*	.53*	.26****

Note. n = number of students
* p< .001 **** p< .05
** p< .01

As the reading correlation coefficients in Table 6 indicate there are fluctuations among progress tests. These fluctuations occur with an abrupt increase and decrease (see Appendix A Fig. 17). What is striking is that there is no correlation between progress test 1, progress test 3, progress test 6 and the end of course assessment test and they do not have statistical significance. On the other hand, progress test 2 and progress test 4 have weak positive correlation with a significance at .01 level. r= .41 implies a moderate positive relationship between progress test 5 and the end of course assessment test. To sum up, it is difficult to say anything about 2-year students' performance on reading section.

Listening section begins with no correlation, followed by moderate positive correlation, again comes no correlation. Then, there is a moderate positive correlation, which is followed by a weak positive one. On the last progress test, strong positive correlation is observed but this is on the border line. There are fluctuations (see Appendix A Fig. 18) The statistical significance of r ranges from no significance to .001 level.

The grammar section indicates more consistent fluctuations. They are not as noticeable as the ones in the reading and listening sections (see Appendix A, Fig. 19). The relationships vary between strong positive relationship and moderate positive relationship. All the values are statistically significant at .001 level.

The correlation coefficients in the writing section are very much like those in Group 2 (see Appendix A, Fig. 20). It starts with a moderate positive relationship which is followed by a weak positive relationship. Both of them are statistically significant but this significance is at different levels. The first one is at the .01 level and the latter is at the .05 level.

To sum up, sections on progress tests do not show very strong positive correlations with the end of course assessment test. The highest correlations are found in the grammar section whereas the lowest ones are in the reading section. Therefore, it seems difficult to predict 2-year students' performance on different sections on the end of course assessment test.

Sub-question 2 : Correlation between Each Section on PTs and on ECA in 4-Year Students in Group 2

4-Year Students

Table 7

Correlation between Each Section on the PTs and ECA in 4-Year Students in Group 2

n=54 (4-year students)	READING ECA	LISTENING ECA	GRAMMAR ECA	RITING ECA
PT 1	.36**	.25*****	.56*	
PT 2	.52*	.32***	.67*	
PT 3	.21	.33***	.57*	.55*
PT 4	.37**	.60*	.64*	
PT 5	.46*	.28****	.73*	
PT 6	.14	.54*	.56*	.43*

Note. n = number of students
* p< .001 **** p< .05
** p< . 01 ***** p< .10
*** p< . 02

As illustrated in Table 7, in the reading section r starts at .36. Like in the other groups, it fluctuates. There are sudden and steep decreases (see Appendix A Fig. 21). For instance, on progress test 2 the correlation is .52 whereas on progress test 3 it is .21. In addition, on progress test 6, the correlation again goes down from .46 to .14, which is a sharp decline.

Progress test 1, 3, 4, and 6 have weak positive correlation with the end of course assessment test. Both progress test 1 and progress test 4 are statistically significant at the .01 level whereas progress test 3 and progress test 6 have no statistical significance. On the other hand, progress test 2 and progress test 5 have moderate positive correlation with the end of course assessment test. Both are significant at the .001 level.

In the listening section, r starts at .25 value. Although the correlation coefficient fluctuates, there is a gradual increase in general(see Appendix A Fig. 22).

There is weak positive correlation between progress test 1,progress test 2,progress test 3,progress test 5 and the end of course assessment test. Progress test 1 is statistically significant at .10 while progress test 2 and progress test 3 are significant at the .02 level. Progress test 5 has a significance at .05 level. The correlation between progress test 4 and the end of course assessment test is strong, positive and statistically significant at the .001 level. Finally, progress test 6 has moderate positive correlation with the end of course assessment test. It is at the .05 level.

As far as grammar section is concerned, there is slight fluctuation between correlation coefficients. More specifically, there is a sustained increase except

progress test 6 (see Appendix A Fig. 23). The correlation coefficient of progress test 6drops from .73 to .56. The correlation coefficient of progress test 1 starts at .56, which is twice as high as the one in listening. Progress test 1, 3, and 6 have moderate positive correlation whereas progress test 2, 4 and 5 have strong positive correlation with the end of course assessment test. All of them are statistically significant at the .001 level.

In the writing section, the correlation coefficient drops from .55 to .43. There is a decrease(see Appendix A Fig. 24). The correlation between progress test 3 and the end of course assessment test is moderately positive and statistically significant at the .001 level whereas the one between progress test 6 and the end of course assessment test is weak positive and statistically significant at the .01 level.

In conclusion, there is fluctuation in all sections. It is interesting to observe that grammar is the section which has the highest correlation on all progress tests when compared to reading, listening, and writing sections. Therefore, estimating students' performance on grammar section on the end of course assessment test is easier than the others. Listening is the only section in which the correlation increased on progress test 6.

CHAPTER 5 CONCLUSIONS

Introduction

This study intended to investigate both the extent to which progress tests helped identify students' overall performance on the end of course assessment test and the extent to which the scores of each section on progress test were indicative of similar performance on the end of course assessment test at the foundation level at Bilkent University School of English Language concerning 2-year and 4-year students.

All the students studying at 2+14 week and 16 week foundation courses participated in the study. However, the ones who missed at least one progress test or the end of course assessment test and the ones who were repeating the foundation level were excluded from the study in order not to affect the reliability of the study. Therefore, the total number of subjects dropped from 348 to 223. 98 of these students were from 2+14 week foundation course and 125 of them were from 16 week foundation course. The researcher also divided the groups according to the length of the study in their future departments as 2-year and 4-year students.

The overall scores of six progress tests and one the end of course assessment test were gathered. Since only the total progress test scores,not the breakdown scores

of progress tests, were recorded on the database of BUSEL, the researcher went through 1338 papers to collect the necessary data except the end of course assessment test. The break-downs in this test are recorded by the teachers after scoring is over.

Next, both the correlation between overall scores on progress tests and on the end of course assessment test, and correlation between each section on progress tests and on the end of course assessment test were calculated. These were analyzed on the basis of Group 1,Group 2 , 2-year and 4-year students groups. Finally, the central tendency and dispersion were computed for overall scores and break-downs on each progress test and on the end of course assessment test.

General Results

The results are discussed in the order of research questions and the data presented in Chapter 4.

Research question 1: Correlation of Overall Scores on PTs and ECA in Group 1 and Group 2

In this study, the researcher first questioned the extent to which the overall scores of progress tests were indicative of foundation level students performance on the end of course assessment test concerning Group 1 , and Group 2.

Group 1 (Overall Scores of Students)

It is found out that using the criteria suggested by Fitz~Gibbon and Morris (1987) Group 1, consisting of all 4-year students, shows a strong positive correlation on progress test 3, 4, 5 and 6 with the end of course assessment test. In other words, students show similar performance on both these progress tests and on the end of course assessment test. The statistical significance of all the values were significant at $p < .001$. The highest correlation is between progress test 4 and the end of course assessment test; the lowest is between progress test 1 and the end of course assessment test.

Group 2 (Overall Scores of Students)

Group 2 includes of both 2-year and 4-year students. This group has very strong positive correlations with the end of course assessment test. All the values are statistically significant at the .001 level. Progress test 5 has the highest correlation and progress test 1 has the lowest correlation with the end of course assessment test.

When Group 1 and Group 2 are compared, it is possible to conclude that both groups show similar patterns, but the degree of correlation is a bit different. In both groups, the correlation between progress test 6 and the end of course assessment test

drops. Group 1 starts at a higher r value on progress test 1 than in Group 2 although Group 2 scored better than Group 1 on the placement test which was administered 5 weeks before progress test 1. To sum up, it seems easier to estimate Group 1 students' overall scores on the end of course assessment test than Group 2 students'.

Sub-question 1: Correlation of Overall Scores on PTs and ECA in 2-year Students Group in Group 2

2-year Students

There are 2-year students only in Group 2. What they show is that they have strong positive correlation between progress test 2, 3, 4,5 and the end of course assessment test. All r values have significance at the .001 level. The highest correlation is between progress test 5 and the end of course assessment test whereas the lowest correlation is between progress test 1 and the end of course assessment test.

Sub-question 2: Correlation of Overall Scores on PTs and ECA in 4-year Students in Group 2

4-year Students

These students have very strong positive correlation between progress test 4, 5 and the end of course assessment test. All the correlations are statistically significant at $p < .001$. Progress test 5 has the highest

correlation and progress test 1 has the lowest correlation with the end of course assessment test.

If 2-year and 4-year students are compared, 4-year students have higher correlations between progress tests and the end of course assessment test than 2-year students do. It can be claimed that it is easier to estimate 4-year students overall scores on the end of course assessment test than 2-year students'.

General Conclusions

In conclusion, it is surprising to find out that the correlation between progress test 6 and the end of course assessment test drops in all groups, that is to say, Group 1, Group 2, 2-year students and 4-year students groups. The highest correlations are between progress test 5 and the end of course assessment test again in all groups. The lowest correlations are between progress test 1 and the end of course assessment test. Group 1 shows the highest correlations of all the groups whereas 2-year students show the lowest ones. Making estimations about students' overall performance is easier in Group 1 than the rest. This is followed by 4-year students in Group 2. Then comes Group 2, the whole group. The most difficult one is 2-year students in Group 2.

Research question 2:Correlation between each Section on the PTs and ECA in Group 1 and Group 2

In the second place, the researcher questioned the extent to which scores of each section on progress tests were indicative of similar performance on the end of course assessment test on the basis of Group 1 and Group 2, 2-year and 4-year students.

Group 1(The Breakdown of Scores)

When each section on progress tests is correlated with the one on the end of course assessment test, it is found that it is not easy to say how students might perform as far as the sections on the end of course assessment test are concerned because correlation coefficients are not high enough.

In the reading section, there is a gradual increase although it is not very high. Progress test 4, 5, and 6 have moderate positive correlations with the end of course assessment test, and the r values are statistically significant at the .001 level. Progress test 1, 2, and 3 have weak positive and statistically significant correlations with the end of course assessment test. This significance range from .01 to .05 level.

In the listening section, the correlation coefficients are higher than the ones in the reading

section. Progress test 4 and progress test 6 have strong positive correlation with the end of course assessment test. The highest correlation is between progress test 4 and the end of course assessment test whereas the lowest one is between progress test 2 and the end of course assessment test. All r values are statistically significant at $p < .001$ level except r on progress test 2. It is statistically significant at $p < .01$ level.

In the grammar section, all progress tests have strong positive correlation with the end of course assessment test except progress test 4 and progress test 6. They have moderate positive correlation with the end of course assessment test. All the correlations show a statistical significance at the .001 level.

As far as the writing section is concerned, both progress tests have moderate positive correlation with the end of course assessment test. They are statistically significant at the .001 level.

Finally, it is surprising to observe that the r value drops in progress test 6 in all the sections with the exception of the listening section. The r value rises in the listening on progress test 6. It is not easy to make estimations about students' performance as far as sections on the end of course assessment test are concerned.

Group 2(The Breakdown of Scores)

This group reveals a moderate positive correlation between only progress test 5 and the end of course assessment test. The other progress tests have either weak positive or almost no correlation with the end of course assessment test. The statistically significance varies between .001 and no significance. Progress test 6 has no statistical significance. The highest correlation is between progress test 5 and the end of course assessment test while the lowest correlation is between progress test 6 and the end of course assessment test. The correlation drops on progress test 6.

In the listening section, progress test 4 and progress test 6 have moderate positive correlation with the end of course assessment test whereas the others have weak positive correlation with the end of course assessment test. The statistical significance is at $p < .001$ level on progress test 2, 4, and 6. On progress test 3 and 5, this significance is at $p < .01$ level and on progress test 1 it is at $p < .05$. Progress test 6 has the highest and progress test 1 has the lowest correlation with the end of course assessment test. Unlike the other sections, the correlation increases on progress test 6.

As far as the grammar section is taken into consideration, progress test 2 and 5 have strong positive

correlation with the end of course assessment test. All the others have moderate positive correlation with the end of course assessment test. All the correlations are statistically significant at the .001 level. the highest correlation occurs between progress test 5 and the end of course assessment test while the lowest correlation is between progress test 1 and the end of course assessment test. The correlation decreases on progress test 6.

In the writing section, there are only two values as it is tested on progress test 3 and 6. The one on progress test 3 has moderate positive correlation and the other has weak positive correlation with the end of course assessment test. The statistical significance is at p< 001 level on both of them. The correlation drops on progress test 6 again.

To summarize, the correlation coefficients in each section fluctuate. The highest correlations occur in the grammar section and the lowest ones are in the reading section. In general, there is a tendency to decline in each section except the listening. The lowest correlations are on progress test 1 except the reading section because the lowest correlations in the reading section are on progress test 6.

Sub-question 1: Correlation between Each Section on PTs and ECA in 2-year Students in Group 2

2-year Students(The Breakdown of Scores)

Progress test 5 is the only one which has moderate positive correlation with the end of course assessment test whereas the other progress tests have either weak positive or no correlation with the end of course assessment test. The correlation coefficients on progress test 2, 4, and 5 are statistically significant at $p<.001$. Progress test 3 is significant at the .10 level while progress test 1 and progress test 6 are not statistically significant. Progress test 5 has the highest and progress test 1 has the lowest correlation with the end of course assessment test.

In the listening section, progress test 6 shows strong positive correlation with the end of course assessment test. The others range from moderate positive to no correlation with the end of course assessment test. Progress test 2, 4, and 6 are significant at the .001 level and progress test 5 is at the .05 level whereas progress test 1 and 3 have no statistical significance. Progress test 6 has the highest and progress test 1 and 3 have the lowest correlation with the end of course assessment test. The correlation rises on progress test 6.

Grammar section has higher correlations than the ones in the other sections. The values vary from .60 to .40. All of them are statistically significant. The correlation drops on progress test 6.

In the writing section, progress test 3 has moderate positive correlation while progress test 6 has weak positive correlation with the end of course assessment test. Progress test 3 is significant at $p < .001$ and progress test 6 as significant at $p < .05$.

Finally, grammar section has higher correlations than the other sections do. There is fluctuation in all sections. The correlation decreases on progress test in all sections except the listening.

Sub-question 2: Correlation between Each Section on PTs and ECA in 4-year Students in Group 2

4-year Students(The Breakdown of Scores)

This group has moderate positive correlation between the reading sections on progress test 2, 5 and on the end of course assessment test. The other correlations are either weak positive or no correlation. Progress test 2 and 5 are statistically significant at the .001; progress test 1 and 4 are at the .01; progress test 3 and 6 are not statistically significant. Progress test 2 has the highest and progress test 6 has the lowest correlation

with the end of course assessment test. The correlation goes down on progress test 6.

In the listening section, progress test 4 has strong positive correlation with the end of course assessment test. The others have either w or moderate positive correlation with the end of course assessment test. The statistical significance varies between .10 and .001 levels. The highest correlation is on progress test 4; the lowest one is on progress test 6. There is a decline in the correlation on progress test 6.

The grammar section reveals either strong or moderate positive correlations with the end of course assessment test. The r values are significant at $p < .001$. Progress test 5 has the highest and progress test 1 and 2 have the lowest correlation with the end of course assessment test. The correlation goes down on progress test 6.

As far as the writing section is concerned, there is moderate positive between progress test 3, 6 and the end of course assessment test. Both of the values have statistical significance at the .001 level. On progress test 6, r falls.

General Conclusions

To conclude, what is striking is that in all groups although the correlation goes down in all sections on

progress test 6, it rises only in the listening section. The highest correlation is indicated in the grammar section in Group 1 while the lowest is in the reading section in 2-year students group. In general, Group 1 has the highest correlations and 2-year students have the lowest ones.

It is not easy to make estimations about students' performance in each section on the end of course assessment test since the correlations are not as high as those in overall scores estimated for the first research question.

Discussion

In this research study, the researcher first investigated the degree of relationship between overall scores on progress tests and on the end of course assessment test. Then, she focused on each section on progress tests and on the end of course assessment test. The data in this study were analyzed using Pearson Product Moment Correlation Coefficient since the data were random, not in a rank as recommended by Brown (1996). In the interpretation stage, the criteria suggested by Fitz~Gibbon and Harris (1987) were used to discuss the strength of the correlation.

During the analysis, it became clear that overall scores on progress tests have higher correlations with

the end of course assessment test than scores of each section on progress tests with corresponding sections on the end of course assessment test. In addition, 2-year students show the lowest correlations both on overall scores and on sections. There might be several factors influencing the correlations.

The first factor might be attributable to the tests and test items. As Brown (1996) stated if the type of items chosen for the test were new to some of the students, this might have influenced the scores. For example, progress test 1 is the first test which the students take. They are not familiar with the test items. Therefore, they might have scored less well on this progress test. Next, if only a small number of items were used in the tests, this might have caused low correlations (Brown, 1996). When a large number of items of are used, the measurement is more accurate. Gronlund (1985) states that " a test is only a sample of the many questions that might be asked. If a test is too short to provide a representative sample of the performance we are interested in, its validity will suffer accordingly"(p. 80). For example, in each progress test, there are limited number of items whereas in the end of course assessment test this number is larger. In addition, the quality of the items might be another reason (Brown,

1996). Item analysis is performed at BUSEL and, where necessary, problematic items are excluded and extra points are added to the overall scores.

The second factor might be due to scoring procedures. For instance, at BUSEL the writing section on progress tests is scored only by the class teacher after the standardisation session. On the other hand, the writing section on the end of course assessment test is checked by two other teachers who do not teach this class. These teachers do not know the students whose papers they score on the end of course assessment test. Therefore, there is less subjectivity in scoring the end of course assessment test. On the other hand, analytic criteria are used to score the writing section on progress tests and the end of course assessment test . Organisation, content, grammar and vocabulary are the sections of these criteria on progress test 3 and 6 whereas there is an additional section on the end of course assessment test, which is mechanics. In this part, punctuation, spelling and legibility of students' handwriting are taken into consideration. On progress test 3 letters are used to give grades. Then, they are converted to numbers. If students score less than "B" in grammar and vocabulary sections, their scores automatically drop. On progress test 6 and the end of

course assessment test numbers are used in scoring. Grammar and vocabulary parts of the criteria do not affect the score as it does on progress test 3. If students score well on other parts like organisation and content, this affects their score positively.

The third factor can be attributed to the examinees. In the first place, test taking strategies affect students' performance. If students do not know how to handle the question or spend too much time on some sections or questions, they might run out of time. Therefore, they might not have time to answer some questions. For instance, while going through the exam papers to record the break-downs, the researcher noticed that some students did not write anything in the writing section or wrote just a few sentences. Since these students did not fulfil the required word limit, their papers were not marked. Consequently, these students scored "0" on the test(s), which might have influenced the scores and correlations. What is more, the characteristics of the examinees can have important affect on the scores. At BUSEL, 2-year students are vocational school students. They do not score as high as 4-year students, who are faculty students, on the University Entrance Exam. Their performance might really be poor not only in high school but also in the

university as well. As Lazaraton and Hatch (1991) stated, "the presence of extremely high and extremely low scores on a variable" might affect the correlation coefficient. For instance, some of the 4-year students are scholarship students. These students are the ones who scored in the top 100 on the University Entrance Exam. There is also a big difference in terms of study habits. 2-year students generally have either poor or no study habits. Of course, this can affect the performance very much. The last but not the least, students background might influence the scores. Students at foundation level generally like grammar-based activities. They do not give as much importance to the other skills as they do to grammar.

The fourth factor might be due to administration time. Progress tests at BUSEL are administered on Monday afternoons. Administering tests after lunch break might lower the performance. In addition, Monday is the day on which people are stressed very much since it is the first day following the weekend. On the other hand, there are times when the tests are given after long holidays or breaks or with long time intervals. Progress test 6 was administered on January, 4, which is just after the New Year break. Students had a 4-day break. Furthermore, progress test 3 was given on November, 2, whereas

university as well. As Lazaraton and Hatch (1991) stated, "the presence of extremely high and extremely low scores on a variable" might affect the correlation coefficient. For instance, some of the 4-year students are scholarship students. These students are the ones who scored in the top 100 on the University Entrance Exam. There is also a big difference in terms of study habits. 2-year students generally have either poor or no study habits. Of course, this can affect the performance very much. The last but not the least, students background might influence the scores. Students at foundation level generally like grammar-based activities. They do not give as much importance to the other skills as they do to grammar.

The fourth factor might be due to administration time. Progress tests at BUSEL are administered on Monday afternoons. Administering tests after lunch break might lower the performance. In addition, Monday is the day on which people are stressed very much since it is the first day following the weekend. On the other hand, there are times when the tests are given after long holidays or breaks or with long time intervals. Progress test 6 was administered on January, 4, which is just after the New Year break. Students had a 4-day break. Furthermore, progress test 3 was given on November, 2, whereas

progress test 4 was on December, 12. There are 5 weeks between these two weeks.

The last factor might be due to instruction. Since progress tests function as formative evaluation, students and teachers are provided with continuous feedback concerning successes and failures. "Feedback to [students] reinforces successful learning and identifies the learning errors that need correction. Feedback to … teacher[s] provides information for modifying instruction and prescribing group and individual remedial work" (Gronlund, 1985, p. 12). At BUSEL , after the feedback provided to the teachers and students, necessary action is taken to improve learning and teaching. For example, students are supplied with tutorials. The students in small groups have extra classes covering the skills which they are weak at so that these weaknesses are treated and the success level improves. Furthermore, to improve the instruction, where necessary, supplementary materials are provided to the teachers. All these affect students' performance.

In conclusion, different factors might have influenced the scores. This difference might result in different correlations both in overall scores and in sections.

Limitations

This study is concerned with both the degree of relationship between overall scores on progress tests and on the end of course assessment test, and the degree of relationship between each section on progress tests and on the end of course assessment test on the basis of Group 1 and Group 2, and 2-year and 4-year students. This study is limited to only foundation level. The other levels such as fou-0, intermediate, upper-intermediate, and pre-faculty have been kept out of the scope of the study because of the short length of the study. In addition, 4-year tourism students are categorized in 2-year students group.

Furthermore, the focus of the study is confined to the degree of relationship between progress tests and the end of course assessment test. It is not concerned with the content validity of these tests since content of these tests are assured by the Curriculum and Testing Unit through test maps.

The findings of the study cannot be generalized to the whole foundation levels or to the other levels.

Implications for Further Practice and Research

The research does not imply that these findings should not be used, only that they are applicable to 1998

registered students at 2+14 week and 16-week foundation level courses.

The findings discussed here should be of some help to the Curriculum and Testing Unit, administrators, and the teachers who are interested in this field so that they might use them for different purposes.

First of all, it is possible to carry out further research on the same topic in different levels, or on the item analysis of either these tests or the others with respect to 2-year and 4-year students so that, if needed, necessary changes might be made on the course schedule. In other words, much more time might be spent on this problematic skills/topics/points.

Second of all, testers might go over the sections which have either low or no correlation to identify where the problems come from.

Third, administrators might observe 2-year and 4-year students' performance more closely so that they can either change the length of the courses like 2+14 week sample or get different exams prepared for 2-year students or provide 2-year students with more support like more tutorials or more experienced teachers or fever number of students in classes, that is to say, the number of students in classes can be decreased, 10 or 12 students in one class. What is more, administrators can

also focus on the departments. As a consequence, they can change the exit level for some departments. For example, intermediate might be the exit level for some departments or for 2-year students.

References:

American Psycological Association. (1985). Standards for educational and psychological testing. Washington, D. C.: American Psycological Association.

Anastasi, A. (1961). Psychological testing. London: MacMillan.

Bachman, L. F.(1990). Fundamental considerations in language testing. Oxford: Oxford University Press.

Bachman, L. F. & Adrian, S.P. (1996). Language testing in practice. Oxford: Oxford University Press.

Bachman, L. F. & Adrian, S.P. (1998). Language testing in practice. Oxford: Oxford University Press

Bloom, B.S., Hastings, J. T. & Madaus, G.F. (1971). Handbook on formative ans summative evaluation of students learning. New York: McGraw-Hill Book Company.

Brown, J. D. (1996). Testing in language programs. New Jersey: Prentice Hall.

Carroll, B. J., & Hall, P.J. (1985). Make your own language tests. New York: Pergamon Press.

Cronbach, L. (1964). Essentials of psychological testing. New York: Harper and Row.

Davies, A. (1984). Communicative language testing. In Proceedings of the seminar, testing English beyond the high school. Istanbul: Bogazici University Publications.

Ebel, R. L. (1980). Evaluation of students: Implications for effective teaching. Educational Evaluation and policy Analysis 2, 47-51.

Finocchiaro, M. & Sako, S. (1983). Foreign Language Testing. New York: Regents Publishing Company.

Fitz~Gibbon, C.T. & Morris, L. L. (1987). How to analyze data. California: the Regents of the University of California.

Gronlund, N. E. (1985). Measurement and evaluation in teaching. New York: Macmillan Publishing Company.

Harlen, W. (1978). Evaluation and the teacher's role. London: Macmillan educational Schools Council Research Studies.

Harris, D.P. (1969). Testing English as a second language. New York: McGraw- Hill Book Company.

Heaton, J.B. (1988). Writing English language tests. Essex: Longman.

Henning, G. (1987). A guide to language testing. Cambridge: Newbury House publishers.

Hills, J.R. (1976). Measurement and evaluation in the classroom. Ohio: A Bell and Howell Company.

Hughes, A. (1997). Testing for language teachers. New York: Cambridge University Press.

Madsen, H.S. (1983). Techniques in testing. New york: Oxford University Press.

Natriello, G. Dornbusch, S. M. (1984). Teacher evaluative standards and student effort. New York: Longman.

Oosterhof, A. (1994). Classroom applications of educational measurement. New York: Macmillan Publishing Company.

Spolsky, B. (1988). Language testing-The problems of validation. In L. Palmer & B. Spolsky, (Eds.), Papers on Language Testing. Washington D.C.

Valette, R. M. (1977). Modern language testing.New York: Harcourt Brace Jovanovich.

APPENDIX A

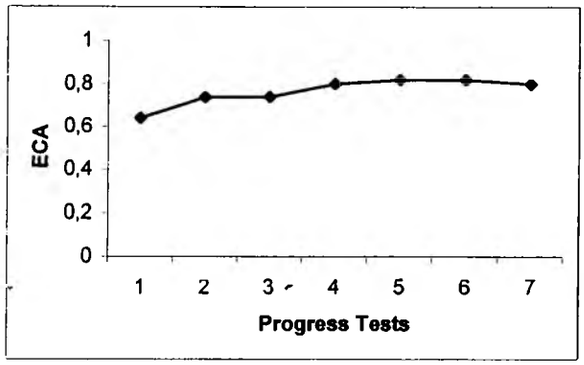

Fig. 5: Correlation between Overall Scores of PTs and of ECA in Group 1

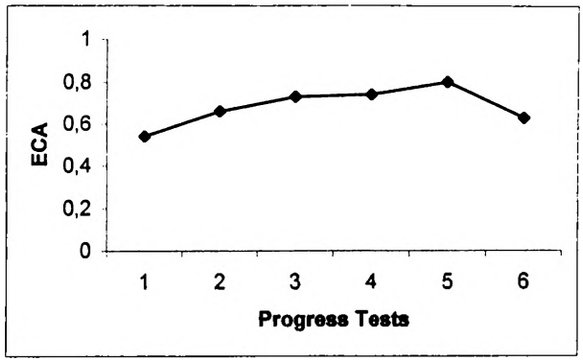

Fig. 6: Correlation between Overall Scores of PTs and of ECA in Group 2

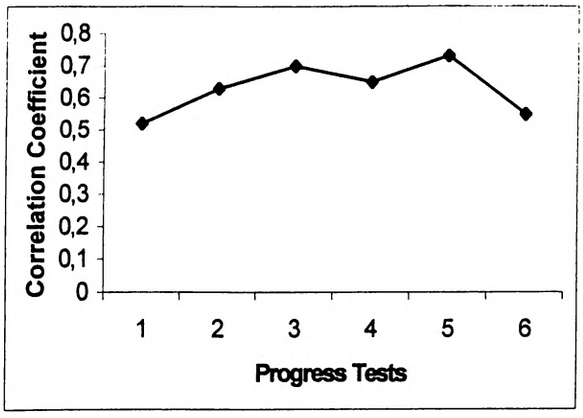

Fig. 7:Correlation between Overall Scores of 2-Year Students in Group 2 on PTs and on ECA

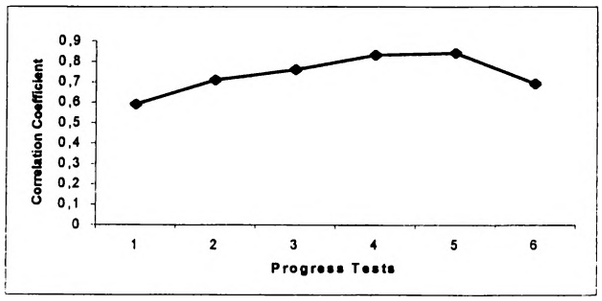

Fig. 8: Correlation between Overall Scores of 4-Year Students in Group 2 on PTs and ECA

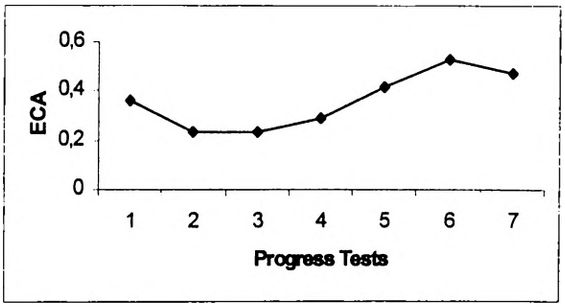

Fig. 9: Correlation between Reading Section on PTs and ECA in Group 1

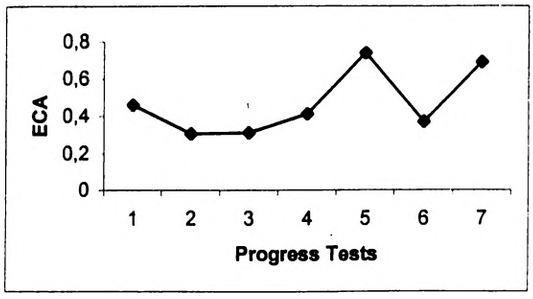

Fig. 10: Correlation between Listening Section on PTs and on ECA in Group 1

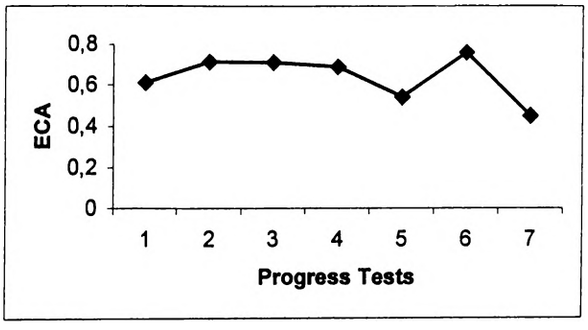

Fig.11 : Correlation between Grammar Section on PTs and on ECA in Group 1

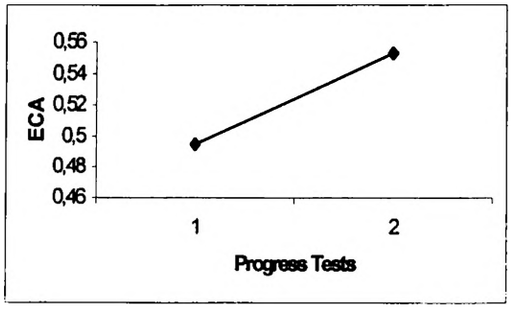

Fig. 12: Correlation between Writing Section on PTs and on ECA in Group 1

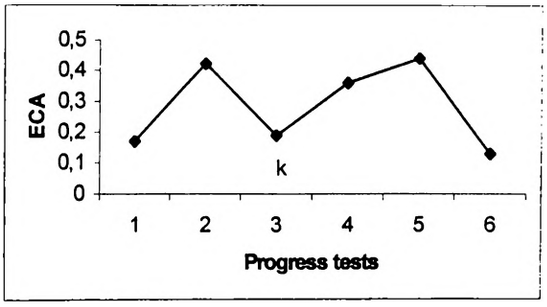

Fig. 13: Correlation between Reading Section on PTs and on ECA in Group 2

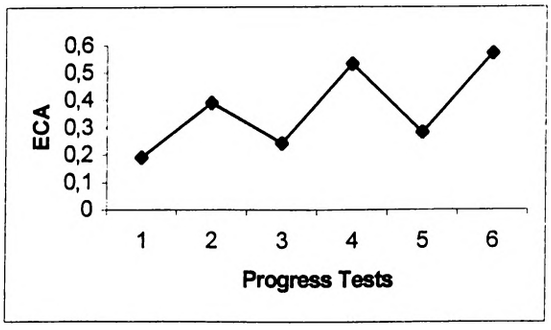

Fig. 14: Correlation between Listening Section on PTs and on ECA in Group 2

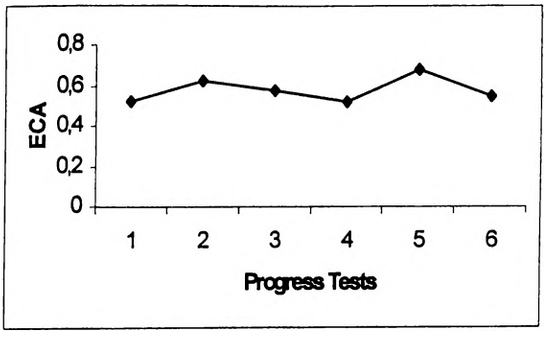

Fig. 15: Correlation between Grammar Section on PTs and on ECA in Group 2

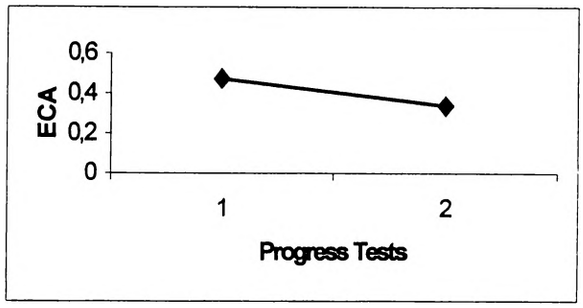

Fig.16: Correlation between Writing Section on PTs and on ECA in Group 2

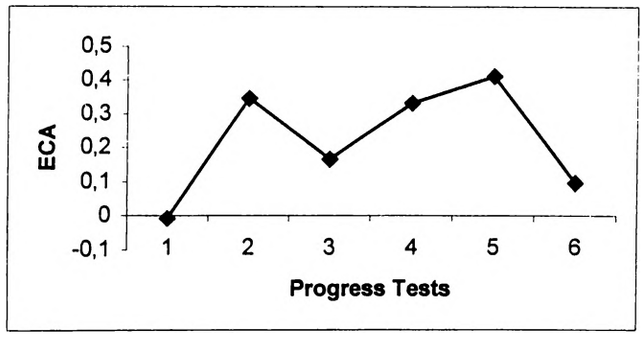

Fig. 17: Correlation between Reading Section on PTs and ECA in 2-Year Students in Group 2

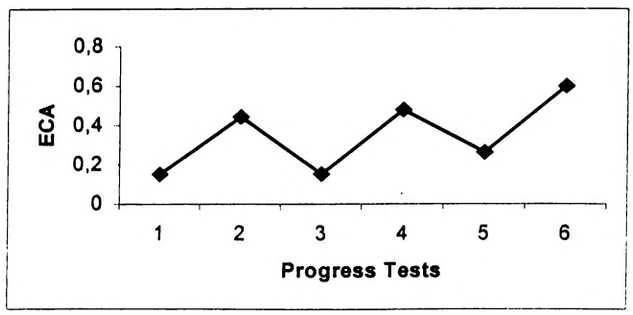

Fig. 18: Correlation between Listening Section on PTs and on ECA in 2-Year Students Group 2

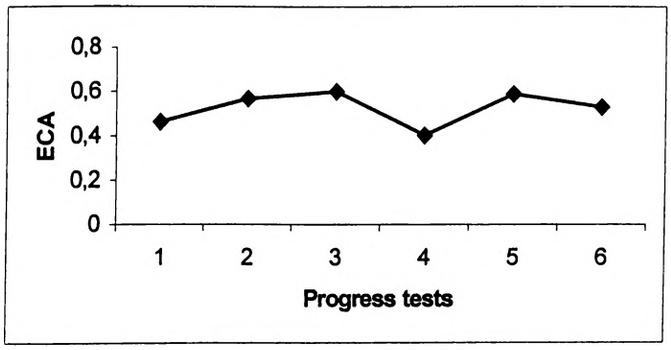

Fig.19: Correlation between Grammar Section on PTs and on ECA in 2-Year Students in Group 2

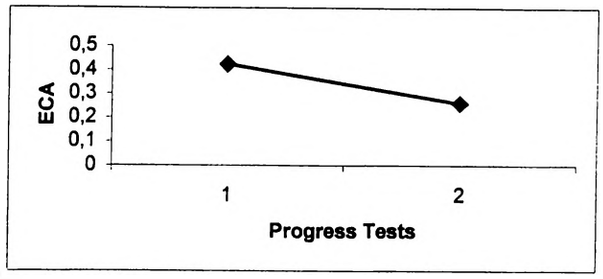

Fig. 20: Correlation between Writing Section on PTs and on ECA in 2-Year Students in Group 2

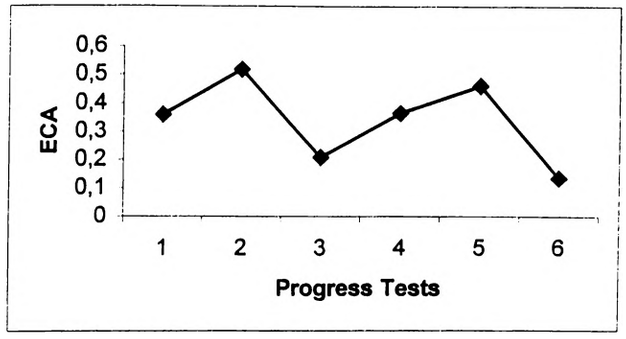

Fig. 21: Correlation between Reading Section on PTs and ECA in 4-Year Students in Group 2

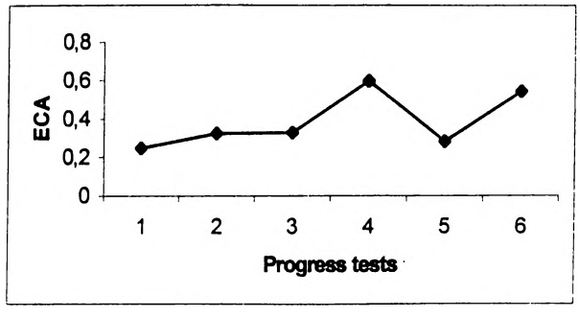

Fig.22: Correlation between Listening Section on PTs and ECA in 4-Year Students in Group 2

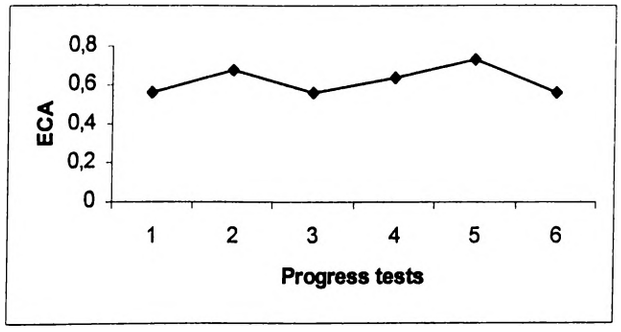

Fig.23: Correlation between Grammar Section on PTs and ECA in 4-Year Students in Group 2

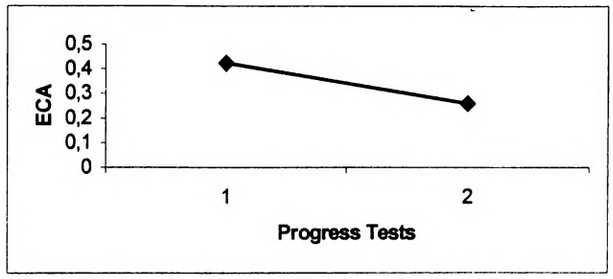

Fig. 24: Correlation between Writing Section on PTs and on ECA in 4-year Students in Group 2

APPENDIX B

Correlation between Overall Scores of PTs and ECA (Vocabulary & Writing not Included)

Group 1

Table 8
Correlation between Overall Scores of PTs and ECA in Group 1 (Vocabulary & Writing not Included)

	ECA	ECA (no vocab and writing)
PRT 1	0,640827	0,648756
PRT 2	0,736953	0,750612
PRT 3	0,79567	0,754513
PRT 4	0,820495	0,802426
PRT 5	0,821556	0,807521
PRT 6	0,799148	0,741823

Group 2

Table 9
Correlation between Overall Scores of PTs and ECA in Group 2 (Vocabulary & Writing not Included)

	ECA	ECA (no vocab and writing)
PRT 1	0,540193	0,488266
PRT 2	0,662556	0,651406
PRT 3	0,732253	0,624089
PRT 4	0,743413	0,721222
PRT 5	0,799406	0,738955
PRT 6	0,628128	0,550822

Table 10
Correlation between Overall Scores of PTs and ECA in 2-year Students in Group 2 (Vocabulary & Writing not Included)

	ECA	ECA (no vocab and writing)
PRT 1	0,515198	0,471318
PRT 2	0,631167	0,639993
PRT 3	0,695752	0,631991
PRT 4	0,651495	0,658737
PRT 5	0,734012	0,659631
PRT 6	0,545071	0,526419

Table 11
Correlation between Overall Scores of PTs and ECA in 4-year Students in Group 2 (Vocabulary & Writing not Included)

	ECA	ECA (no vocab and writing)
PRT 1	0,588903	0,520047
PRT 2	0,705108	0,675717
PRT 3	0,757036	0,60048
PRT 4	0,833434	0,779258
PRT 5	0,844435	0,785224
PRT 6	0,689504	0,557625

APPENDIX C

The Measures of Central Tendency and Dispersion of Overall Scores and Each Section on ECA and PTs

Group 1

Table 12

The Central Tendency and Dispersion of ECA in Group 1

ECA	READ	LISTEN	GRAM	WRITE	OVERALL
Mean	11,35714	8,826531	33,41837	12,3449	65,94694
Standard Error	0,30469	0,322967	0,887523	0,44604	1,688943
Median	12	8	33,5	12	67
Mode	14	8	32	12	52,9
Standard Deviation	3,016279	3,197214	8,786029	4,41558	16,71968
Sample Variance	9,097938	10,22218	77,1943	19,4973	279,5477
Skewness	-0,41533	0,128489	-0,30758	-1,0477	-0,27089
Range	10	12,5	35	20	67,1
Minimum	5	2,5	14	0	27,5
Maximum	15	15	49	20	94,6

Table 13

The Central Tendency and Dispersion of PT 1 in Group 1

PT 1	READ	LISTEN	GRAM	VOCAB	OVERALL
Mean	4,09183	3,73469	13,1122	14,5051	17,7520
Standard Error	0,111496	0,107132	0,18051	0,07270	0,167117
Median	4,5	4	14	15	18
Mode	5	4	14	15	19
Standard Deviation	1,103759	1,060549	1,78702	0,71973	1,654374
Sample Variance	1,218283	1,124763	3,19345	0,51801	2,736955
Skewness	-0,98326	-0,55876	-1,0353	-1,7541	-0,72545
Range	4	4	8	3	7,2
Minimum	1	1	7	12	12,8
Maximum	5	5	15	15	20

Table 14

The Central Tendency and Dispersion of PT 2 in Group 1

PT 2	READ	LISTEN	GRAM	OVERALL
Mean	4,785714	3,877551	15,7398	24,41327
Standard Error	0,072886	0,115372	0,290844	0,384314
Median	5	4	16,25	25
Mode	5	5	18	28
Standard Deviation	0,721539	1,142121	2,879209	3,804518
Sample Variance	0,520619	1,304439	8,289843	14,47436
Skewness	-3,51799	-0,85714	-0,71273	-0,90624
Range	4	5	12	17,5
Minimum	1	0	8	12,5
Maximum	5	5	20	30

Table 15

The Central Tendency and Dispersion of PT 3 in Group 1

PT 3	READ	LISTEN	GRAM	WRITE	OVERALL
Mean	4,311224	4,239796	19,85204	3,520408	32,0051
Standard Error	0,085048	0,062485	0,356052	0,118931	0,51126
Median	4,75	4,5	20,5	4	33,75
Mode	5	4	24	4	34
Standard Deviation	0,841927	0,618568	3,524735	1,17736	5,061221
Sample Variance	0,708842	0,382627	12,42376	1,386177	25,61595
Skewness	-1,18197	-0,95272	-0,93623	-0,64957	-0,81656
Range	3	3	16,5	5	20,5
Minimum	2	2	8,5	0	18,5
Maximum	5	5	25	5	39

Table 16

The Central Tendency and Dispersion of PT 4 in Group 1

PT 4	READ	LISTEN	GRAM	VOCAB	OVERALL
Mean	4,045918	3,204082	5,153061	3,663265	16,06122
Standard Error	0,06768	0,145382	0,075512	0,059707	0,251365
Median	4	3,5	5,5	4	16,5
Mode	4,5	5	5,5	4	17,5
Standard Deviation	0,670001	1,439209	0,747533	0,591065	2,488391
Sample Variance	0,448901	2,071323	0,558805	0,349358	6,192089
Skewness	-0,35456	-0,34542	-0,95714	0,062455	-0,37811
Range	2,5	4,5	3,5	4,5	9,5
Minimum	2,5	0,5	2,5	2	10,5
Maximum	5	5	6	6,5	20

Table 17

The Central Tendency and Dispersion of PT 5 in Group 1

PT 5	READ	LISTEN	GRAM	VOCAB	OVERALL
Mean	3,605882	4,076471	7,152941	4,282353	19,08163
Standard Error	0,120716	0,06721	0,263305	0,183997	0,479689
Median	3,5	4	7	4,5	18,5
Mode	3,5	4	7	4,5	17
Standard Deviation	1,112949	0,619648	2,42755	1,696367	4,748681
Sample Variance	1,238655	0,383964	5,892997	2,877661	22,54997
Skewness	-0,80391	-0,68215	0,373339	-0,20708	0,091132
Range	4,5	2,5	9,5	6,5	21
Minimum	0,5	2,5	3	0,5	8
Maximum	5	5	12,5	7	29

Table 18

The Central Tendency and Dispersion of PT 6 in Group 1

PT 6	READ	LISTEN	GRAM	WRITE	VOCAB	OVERALL
Mean	7,102041	5,316327	7,632653	3,513265	4,168367	27,74796
Standard Error	0,162488	0,191524	0,111545	0,12865	0,07639	0,452048
Median	8	5,5	7,5	3,7	4,5	28
Mode	8	5,5	8	4,3	5	29
Standard Deviation	1,608545	1,895991	1,104235	1,273573	0,756225	4,475045
Sample Variance	2,587418	3,594782	1,219335	1,621987	0,571876	20,02603
Skewness	-0,7894	-0,08455	-0,34429	-1,423	-0,88486	-0,08616
Range	8	8	5	5	3	18,5
Minimum	2	1	4,5	0	2	18
Maximum	10	9	9,5	5	5	36,5

Group 2

Table 19

The Central Tendency and Dispersion of ECA in Group 2

ECA	READ	LISTEN	GRAM	WRITE	OVERALL
Mean	11,72	9,568	33,392	13,2224	67,9024
Standard Error	0,22933	0,23516	0,64935	0,27017	1,13651
Median	12	9,5	34	13,4	67,5
Mode	14	8,5	30	13,4	76
Standard Deviation	2,56401	2,62919	7,26002	3,02059	12,7066
Sample Variance	6,57419	6,91267	52,708	9,12401	161,458
Skewness	-0,48498	-0,03192	-0,09536	-0,43814	-0,07629
Range	9	12	30	20	54,1
Minimum	6	3	17	0	41,5
Maximum	15	15	47	20	95,6

Table 20

The Central Tendency and Dispersion of PT 1 in Group 2

PT 1	READ	LISTEN	GRAM	VOCAB	OVERALL
Mean	4,88	5,584	4,838	2,95967	18,2472
Standard Error	0,03331	0,062922	0,06673	0,02534	0,127411
Median	5	6	5	3	18,5
Mode	5	6	5	3	19
Standard Deviation	0,37243	0,70349	0,74613	0,28224	1,42450
Sample Variance	0,13871	0,494903	0,55671	0,07966	2,029206
Skewness	-3,26679	-2,06947	-0,70839	-9,66823	-1,43651
Range	2	3,5	3,5	3	7,3
Minimum	3	2,5	2,5	0	12,7
Maximum	5	6	6	3	20

Table 21

The Central Tendency and Dispersion of PT 2 in Group 2

PT 2	READ	LISTEN	GRAM	OVERALL
PT 2	READ	LISTEN	GRAM	TOTAL
Mean	6,36065	5,01639	11,34836	22,784
Standard Error	0,07367	0,15615	0,197008	0,31244
Median	7	5	11,5	22,5
Mode	7	6	11,5	20,5
Standard Deviation	0,81378	1,7248	2,176024	3,49327
Sample Variance	0,66224	2,97493	4,73508	12,2029
Skewness	-1,12913	-0,2713	-0,48688	-0,2012
Range	3	7	10	15,5
Minimum	4	1	5	14,5
Maximum	7	8	15	30

Table 22

The Central Tendency and Dispersion of PT 3 in Group 2

PT 3	READ	LISTEN	GRAM	OVERALL
Mean	3,464	2,984	14,428	33,06
Standard Error	0,06960	0,06115	0,159303	0,34643
Median	4	3	14,5	33,5
Mode	4	3	16	36
Standard Deviation	0,77817	0,68372	1,781065	3,87329
Sample Variance	0,60554	0,46748	3,172194	15,0024
Skewness	-1,54809	0,19315	-0,61975	-0,2613
Range	3	4,5	7,5	16,5
Minimum	1	1	9,5	23,5
Maximum	4	5,5	17	40

Table 23

The Central Tendency and Dispersion of PT 4 in Group 2

PT 4	READ	LISTEN	GRAM	Vocab	OVERALL
Mean	3,988	3,984	4,884	3,612	16,448
Standard Error	0,06683	0,08649	0,079542	0,048148	0,181199
Median	4	4	5	4	17
Mode	4	5	5,5	4	18
Standard Deviation	0,74721	0,96707	0,889309	0,538307	2,025865
Sample Variance	0,55832	0,93522	0,790871	0,289774	4,104129
Skewness	-1,43179	-0,9531	-0,73473	-1,12955	-0,45488
Range	5	4	3,5	2	8
Minimum	0	1	2,5	2	11,5
Maximum	5	5	6	4	19,5

Table 24

The Central Tendency and Dispersion of PT 5 in Group 2

PT 5	READ	LISTEN	GRAM	VOCAB	OVERALL
Mean	3,482	4,204	6,86	4,036	18,5976
Standard Error	0,09008	0,05877	0,182421	0,164236	0,367595
Median	3,5	4,5	6,5	4,5	18,5
Mode	4	4,5	6	5,5	19,5
Standard Deviation	1,00712	0,65714	2,039529	1,836213	4,109832
Sample Variance	1,01429	0,43183	4,159677	3,371677	16,89072
Skewness	-0,52053	-1,0312	0,328553	-0,31941	0,204686
Range	4,5	3,5	9,5	6,5	17
Minimum	0,5	1,5	2,5	0,5	10,5
Maximum	5	5	12	7	27,5

Table 25

The Central Tendency and Dispersion of PT 6 in Group 2

PT 6	READ	LISTEN	GRAM	WRITE	VOCAB	OVERALL
Mean	7,044	6,116	7,088	3,2064	3,996	27,512
Standard Error	0,17124	0,16354	0,130268	0,1082	0,086229	0,424359
Median	8	6	7,5	3,3	4,5	27,3
Mode	8	6	8	4	5	27
Standard Deviation	1,91452	1,82846	1,456443	1,2107	0,964064	4,744481
Sample Variance	3,66538	3,34329	2,121226	1,4659	0,929419	22,5101
Skewness	-1,13272	-0,1787	-0,76407	-0,959	-0,77054	-0,03845
Range	10	8	7	5	4	21,7
Minimum	0	1,5	2,5	0	1	17
Maximum	10	9,5	9,5	5	5	38,7

2-year Students in Group 2

Table 26
The Central Tendency and Dispersion of ECA in 2-year Students in Group 2

	ECA
Mean	66,05634
Standard Error	1,321063
Median	66,9
Mode	66,9
Standard Deviation	11,13147
Sample Variance	123,9096
Skewness	-0,14454
Range	51,5
Minimum	41,5
Maximum	93

Table 27
The Central Tendency and Dispersion of PT 1, PT 2 and PT 3 in 2-year Students in Group 2

	PT 1	PT 2	PT 3
Mean	18,30211	22,6831	32,67606
Standard Error	0,15036	0,40111	0,422112
Median	18,5	22,5	33,5
Mode	19,25	20,5	36
Standard Deviation	1,266955	3,379813	3,556783
Sample Variance	1,605174	11,42314	12,6507
Skewness	-1,70667	-0,16102	-0,45843
Range	7,3	15,5	14,5
Minimum	12,7	14,5	24
Maximum	20	30	38,5

Table 28
The Central Tendency and Dispersion of PT 4, PT 5 and PT 6 in 2-year Students in Group 2

	PT 4	PT 5	PT 6
Mean	16,30282	17,72817	28,90741
Standard Error	0,230798	0,433122	0,639585
Median	17	18	29,1
Mode	17,5	15	29,5
Standard Deviation	1,944734	3,649547	4,699974
Sample Variance	3,781992	13,3192	22,08976
Skewness	-0,66424	0,068743	0,160545
Range	8	14,5	19
Minimum	11,5	10,5	19,7
Maximum	19,5	25	38,7

4-year Students in Group 2

Table 29
The Central Tendency and Dispersion of ECA in 4-year Students in Group 2

	ECA
Mean	70,32963
Standard Error	1,941411
Median	71,75
Mode	86,5
Standard Deviation	14,2664
Sample Variance	203,53
Skewness	-0,24544
Range	54,1
Minimum	41,5
Maximum	95,6

Table 30
The Central Tendency and Dispersion of PT 1, PT 2 and PT 3 in 4-year Students in Group 2

	PT 1	PT 2	PT 3
Mean	18,175	22,91667	33,56481
Standard Error	0,220193	0,498732	0,576297
Median	18,875	22,75	33,75
Mode	19	27	35
Standard Deviation	1,61808	3,664915	4,234902
Sample Variance	2,618184	13,4316	17,9344
Skewness	-1,19783	-0,26484	-0,22508
Range	6,75	14,5	16,5
Minimum	13,25	14,5	23,5
Maximum	20	29	40

Table 31
The Central Tendency and Dispersion of PT 4, PT 5 and PT 6 in 4-year Students in Group 2

	PT 4	PT 5	PT 6
Mean	16,63889	19,74074	28,90741
Standard Error	0,290009	0,602194	0,639585
Median	17	19,5	29,1
Mode	18	19,5	29,5
Standard Deviation	2,131119	4,425207	4,699974
Sample Variance	4,541667	19,58246	22,08976
Skewness	-0,30733	0,07799	0,160545
Range	8	15	19
Minimum	11,5	12,5	19,7
Maximum	19,5	27,5	38,7

The Measures of Central Tendency and Dispersion of ECA and PTs
(Vocabulary & Writing not Included)

Group 1

Table 32
The Central Tendency and Dispersion of ECA in Group 1 (Vocabulary & Writing not Included)

	ECA
	ECA out of 80 (No Writing)
Mean	53,60204
Standard Error	1,364274
Median	54,75
Mode	71,5
Standard Deviation	13,50562
Sample Variance	182,4019
Skewness	-0,19614
Range	52,5
Minimum	24
Maximum	76,5

Table 33
The Central Tendency and Dispersion of PT 1, PT 3 and PT 4in Group 1 (Vocabulary & Writing not Included)

	PT 1	PT 3	PT 4
	out of 12,5 (No Vocab)	out of 35 (No Writing)	out of 16 (No Vocab)
Mean	10,52041	28,40306	12,40306
Standard Error	0,153834	0,429041	0,230334
Median	10,75	29	12,5
Mode	11,5	31,5	15
Standard Deviation	1,522881	4,24729	2,280191
Sample Variance	2,319167	18,03948	5,199269
Skewness	-0,87538	-0,93452	-0,27529
Range	7	17,5	8,5
Minimum	5,5	16,5	7,5
Maximum	12,5	34	16

Table 34
The Central Tendency and Dispersion of PT 5, and PT 6 in Group 1 (Vocabulary & Writing not Included)

	PT 5 out of 23 (No Vocab)	PT 6 out of 30 (No Vocab & Writing)
Mean	14,83529	14,83529
Standard Error	0,373246	0,373246
Median	14,5	14,5
Mode	14,5	14,5
Standard Deviation	3,441162	3,441162
Sample Variance	11,8416	11,8416
Skewness	0,100116	0,100116
Range	14,5	14,5
Minimum	7,5	7,5
Maximum	22	22

Group 2

Table 35
The Central Tendency and Dispersion of ECA in Group 2 (Vocabulary & Writing not Included)

	ECA out of 80 (No Writing)
Mean	54,68
Standard Error	0,966126
Median	54,5
Mode	57
Standard Deviation	10,80162
Sample Variance	116,675
Skewness	-0,09281
Range	46
Minimum	31
Maximum	77

Table 36
The Central Tendency and Dispersion of PT 1, PT 3 in Group 2 (Vocabulary & Writing not Included)

	PT 1	PT 3	PT 3
	out of 17 (No Vocab)	Out of 35 (No Vocab)	out of 25 (No Vocab & Writing)
Mean	15,302	27,94	20,876
Standard Error	0,115817	0,347229	0,216173
Median	15,5	28	21
Mode	16,25	23	21
Standard Deviation	1,294878	3,882134	2,416889
Sample Variance	1,67671	15,07097	5,841355
Skewness	-1,34118	-0,19383	-0,43172
Range	6,75	16	10
Minimum	10,25	19	15
Maximum	17	35	25

Table
Table 37
The Central Tendency and Dispersion of PT 4, PT 5 and PT 6 in Group 2 (Vocabulary & Writing not Included)

	PT 4	PT 5	PT 6	PT 6
	out of 16 (No Vocab)	out of 23 (No Vocab)	out of 35 (No Vocab)	out of 30 (No Vocab & Writing)
Mean	12,856	14,546	23,4544	20,248
Standard Error	0,162125	0,250329	0,365971	0,322472
Median	13	14	23,5	20,5
Mode	13,5	12	23,5	19,5
Standard Deviation	1,812608	2,798758	4,091681	3,605341
Sample Variance	3,285548	7,833048	16,74186	12,99848
Skewness	-0,58401	0,314516	-0,24053	-0,28372
Range	8	11,5	19,5	15,5
Minimum	7,5	9	13	12
Maximum	15,5	20,5	32,5	27,5

2-year Students in Group 2

Table 38
The Central Tendency and Dispersion of ECA in 2-year Students in Group 2 (Vocabulary & Writing not Included)

	ECA out of 80 (No Writing)
Mean	53,05634
Standard Error	1,118104
Median	53,5
Mode	49,5
Standard Deviation	9,421309
Sample Variance	88,76107
Skewness	-0,0342
Range	42
Minimum	31
Maximum	73

Table 39
The Central Tendency and Dispersion of PT 1, PT 3 and PT 4 in 2-year Students in Group 2 (Vocabulary & Writing not Included)

	PT 1	PT 3	PT 4
	out of 17 (No Vocab)	out of 25 (No Vocab & Writing)	out of 16 (No Vocab)
Mean	15,35211	21,2963	12,71127
Standard Error	0,128896	0,347578	0,205474
Median	15,5	21,5	13
Mode	16,25	24	13,5
Standard Deviation	1,086095	2,554165	1,731354
Sample Variance	1,179603	6,52376	2,997586
Skewness	-1,01413	-0,56472	-0,62075
Range	5	9,5	7,5
Minimum	12	15,5	8
Maximum	17	25	15,5

Table 40
The Central Tendency and Dispersion of PT 5, and PT 6 in 2-year Students in Group 2 (Vocabulary & Writing not Included)

	PT 5	PT 6
	out of 25 (No Vocab)	out of 25 (No Vocab & Writing)
Mean	13,91901	19,58451
Standard Error	0,284084	0,432575
Median	13,5	19,5
Mode	12	19,5
Standard Deviation	2,393732	3,644944
Sample Variance	5,729955	13,28561
Skewness	0,261639	-0,35708
Range	10,75	13,5
Minimum	9,5	12
Maximum	20,25	25,5

4-year Students in Group 2

Table 41
The Central Tendency and Dispersion of ECA in 4-year Students in Group 2 (Vocabulary & Writing not Included)

	ECA
	out of 80 (No Writing)
Mean	56,81481
Standard Error	1,653105
Median	58,5
Mode	57
Standard Deviation	12,14779
Sample Variance	147,5688
Skewness	-0,35383
Range	46
Minimum	31
Maximum	77

Table 42
The Central Tendency and Dispersion of PT 1, PT 3 and PT 4 in 4-year Students in Group 2 (Vocabulary & Writing not Included)

	PT 1	PT 3	PT 4
	out of 17 (No Vocab)	out of 25 (No Vocab & Writing)	out of 16 (No Vocab)
Mean	15,21759	21,2963	13,0463
Standard Error	0,208469	0,347578	0,260445
Median	15,875	21,5	13,5
Mode	16	24	14
Standard Deviation	1,531929	2,554165	1,913873
Sample Variance	2,346807	6,52376	3,662911
Skewness	-1,35254	-0,56472	-0,62913
Range	6,75	9,5	8
Minimum	10,25	15,5	7,5
Maximum	17	25	15,5

Table 43
The Central Tendency and Dispersion of PT 5, and PT 6 in 4-year Students in Group 2 (Vocabulary & Writing not Included)

	PT 5	PT 6
	out of 25 (No Vocab)	out of 25 (No Vocab & Writing)
Mean	15,37037	21,12037
Standard Error	0,420306	0,461467
Median	15,25	21
Mode	14	23,5
Standard Deviation	3,088605	3,391075
Sample Variance	9,539483	11,49939
Skewness	0,058845	-0,10303
Range	11,5	14,5
Minimum	9	13
Maximum	20,5	27,5